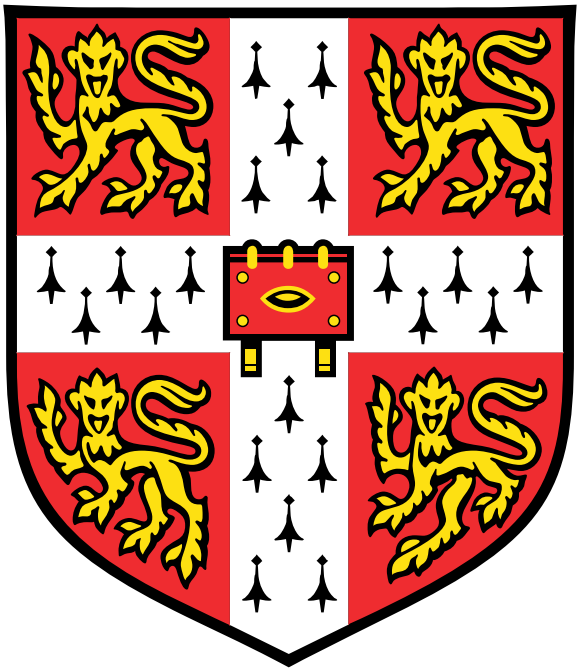
The Machine–Organism Distinction



Tyler David Price Brunet

Supervisor: Dr. T. Lewens

Department of History and Philosophy of Science
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Clare College

May 2021

For K.F.P & V.P.B

Declaration

This thesis is submitted for the degree of Doctor of Philosophy. This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the Department of History and Philosophy of Science Degree Committee.

Tyler David Price Brunet
May 2021

Abstract

Title: The Machine–Organism Distinction
Author: Tyler David Price Brunet
The idea that analysis of organisms can proceed by distinguishing organisms from machines is common to many areas of philosophy. This thesis argues that our search for a philosophy of organisms should not proceed by defining or relying on a Machine–Organism Distinction (MOD). We are often able to take biological theories that are thought to characterize organisms, such as theories of organismal autonomy and stability, and apply them to machines. I argue that we should not provide an analysis of organisms according to an MOD because there is no distinction available that holds up to scrutiny and evidence. There have been several major attempts to provide an MOD. I divide these in consecutive chapters according to the property of organisms offered as an MOD: teleology (Nicholson 2013), autonomy (Mossio and Moreno 2015), stochasticity (Skillings 2015; Godfrey-Smith 2016) and processual stability (Dupré and Nicholson 2018). I address these major attempts to provide an MOD by showing how each fails to provide an analysis of organisms that distinguishes them from machines. To do this, I examine a diversity of machines and organisms that serve as naturalistic counterexamples. Discoveries in molecular biology and ecology, as well as developments in robotics and biotechnology, show the failure of MODs in contemporary philosophy and biology. Moreover, not only does the MOD consistently fail, but philosophical arguments that rely upon MODs consistently misrepresent organisms themselves. I conclude with the idea that we should consider machines not as external to, or distinguished from, organisms, but as proper objects of biological science.

Acknowledgements

This thesis could not have been written without the support of my partner Ariane Hanemaayer. My deepest gratitude also goes to my friends and family, who kept me sane over the duration of writing. Especially to Adrian Erasmus, Adi Avidani, Eyal Fisher, Mandy Wigdorowitz, Olesya Bondarenko, Azita Chellappoo, T. Y. Willian Wong, Catherine Murray, MacGregor Malloy, Ciddia Brunet, Erik Lankin, Vincent Brunet and Kathryn Price. The thesis would never have begun without the initial support of W. Ford Doolittle, nor without the feedback and inspiration of Austin Booth, Carlos Mariscal and Andrew Inkpen. I am very grateful to Marta Halina for collaborating on a paper developed in part from the fifth chapter of this thesis, to John Dupré for supportive discussions of my ideas for the forth chapter, and to the organizers and attendees of the ISHPSSB in Oslo 2019, for providing an opportunity for me to present a portion of my research. I would also like to thank the Social Sciences and Humanities Research Council of Canada and the Cambridge Commonwealth and European International Trust for funding my studies. Special thanks go to those who inspired the thesis and read and responded to its chapters, to my advisor Jacob Stegenga and supervisor Tim Lewens.

Table of contents

List of figures

Nature and Art, they go their separate ways,
It seems; yet all at once they find each other.
Even I no longer am a foe to either;
Both equally attract me nowadays.

*

Some honest toil's required; then, phase by phase,
When dilligence and wit have worked together
To tie us fast to Art with their good tether,
Nature again may set our hearts ablaze...

—Goethe, *Nature and Art*

Chapter 1

Introduction

1.1 The Machine–Organism Distinction (MOD)

In this thesis I examine ways that machines have been distinguished from organisms, or, attempts to provide a Machine–Organism Distinction (MOD). Despite the apparent desirability of an MOD, there is no such distinction that holds up to scrutiny. There are always borderline cases and blatant violations; for any MOD, none cuts Nature at its joint with Art. I show repeatedly, across topics in philosophy, that providing an MOD is a common logical device and rhetorical move. The following two argument structures are common: Because there is some MOD, some claim or theory T is true; or, Because T is true, there is some MOD. But there is no empirically adequate MOD; at least, none sufficiently strong to support arguments of that sort. I argue against each MOD currently on offer, showing how it does not survive analysis or live up to examples from contemporary biology and engineering.

This thesis is about rejecting distinctions between machines and organisms. To do that, I raise many examples of similarities between machines and organisms, but it is not merely a defence of the idea that machines are similar to organisms. Machines *are* similar to organisms in some ways, different in others, most of which are obvious enough not to deserve defence. No known organism is composed entirely of iron, no known machine collects blue buttons to adorn its nesting site. But these aren't interesting distinctions because nobody thinks that being entirely composed or iron of collecting blue buttons are fundamental to our concepts of machines or organisms. This thesis addresses differences that are thought, often for very good reasons, to *define* or *characterize* organisms and machines. As such, it is a defence of the idea that machines are similar to organisms *in ways thought characteristic of organisms*, or vice versa as the case may be. To argue for this, I supply naturalistic counterexamples, of type I and II, for each MOD discussed: examining only extant and known cases, for each MOD, I show (type I) that there is a machine that possesses a property thought to

characterize organisms, and I show (type II) that there is an organism lacking a property thought characteristic of organisms.

MODs often arise from earnest attempts to pick out some fundamental and important feature of organisms by contrast with machines. This is *prima facie* a good strategy, and a large part of the work of introducing MODs throughout this thesis is done to show just how often this happens in the philosophy of science. Consider one example, a detailed examination of which will appear later (Ch.3). Organisms seem to be peculiarly autonomous things: they can move on their own, seek out food and shelter and evade capture, they can reproduce their own kind, etc. So it seems that autonomy must somehow figure as a fundamental, special or important feature in our analysis of organisms. One *prima facie* good way to justify that some property is fundamental or important for a class is to construct a contrast class. So it seems that justifying the special place of autonomy in our analysis of organisms should show how it contrasts with the non-autonomy of non-organisms. Here is the point where the MOD, unhelpfully and wrongly I argue, comes to serve a logical role in debates about our analysis of organisms. Since machines are non-organisms, presumed members of the contrast class, they must be non-autonomous. On the basis of a good-faith attempt to understand organisms, this provides an MOD: organisms are autonomous and this distinguishes them from machines, which are non-autonomous. However (as I will argue in detail in Ch.3), machines are *not* all non-autonomous, nor are organisms all autonomous. Confronted by cases (type I) of autonomous machines from computer engineering and robotics, and (type II) of non-autonomous organisms drawn from microbiology and biochemistry, the MOD based on autonomy fails as a justification for an analysis of organisms that defines them as fundamentally, specially autonomous.

The MODs addressed in this thesis are drawn from contemporary debates within the philosophy of science, of biology in particular. They are divided here by topic into MODs based, respectively, on biological teleology (Ch.2), autonomy (Ch.3), ontology (Ch.4-5) and evolutionary theory (Ch.6). However, each has historical and conceptual roots deeper within general philosophy. Sometimes, the authors of MOD themselves trace the antecedents of their views back to keystone philosophical ideas. For example, the MODs appearing in conceptions of organizational teleology (Nicholson 2013) and biological autonomy (Moreno and Mossio 2015; Maturana and Varela 1972) trace back to Kant's ideas about the internal grounding of purposiveness in his *Third Critique*. Likewise, contemporary debate surrounding the correct metaphysics of organisms can be divided into two camps. One, dealing with a cluster of views termed New-Mechanist philosophy of science (Machamer et al. 2000; Bechtel 2011; Skillings 2015; Godfrey-Smith 2016), traces back to Descartes' ideas about the dispositions of animal bodies in his *Treaties on Man*. Another, alternative, metaphysics of science termed

Process Ontology (Dupré 2018; 2013; Seibt 2018; Rescher 1996; Dupré and Nicholson 2018) traces itself much farther back, to the metaphysics of change derived from the fragments of Heraclitus. I will note these connections to general philosophy as they arise throughout this thesis. Resolving these background issues on which the MOD currently takes centre-stage is not what this thesis is about. That said, I hope to convince the sceptical reader that the MOD is not an isolated issue, not peculiar to contemporary philosophy of biology, but a pernicious conceptual distinction arising in many different philosophical quarters.

At the end of this thesis, after the work of rooting out MODs is finished, I speculate on the biological relationship between us and our machines. I argue that it is precisely analogous to the relationship between non-human animals and their animal-artefacts, and to the relationship between social organisms and their worker castes. I argue that the proper relationship between machines and organisms is captured by part-whole relationships. Our machines are, for now at least, parts of us in the same way that spider-webs are parts of spiders, and worker bees are parts of their queen or their hive. If machines are to be "distinguished" from organisms at all, it is as parts are distinguished from their wholes. However, even this part-whole relationship can change, I argue, and is not sufficient to support MODs of the sort that have been offered previously in the philosophy of biology.

The MOD is an old issue. However, the following section (§ 1.2) argues that it is worth reconsidering now, given developments in both biology and engineering science. We have discovered a greater diversity of organisms and their ways of life than were known to early modern philosophers, and have developed a greater diversity of machines—so much so that the gaps there once were between them are closing fast. Moreover, I argue that neglect of the diversity in contemporary machines in particular leads to a philosophical methodology that is detrimental to our understanding of organisms. In the best case, our philosophy of the natural world would be supported by a thoroughgoing naturalism that accounted for the rich diversity of both biological and technological things—and in so doing we would see that provision of MODs is a misguided way of analysing organisms.

1.2 Contemporary Diversity In Machines and Organisms

Clocks are untimely. They were very sophisticated machines in the 18th and 19th centuries. Now, by comparison, they are not particularly special. Clocks are not even the most common machines; cars would be better, or light-bulbs, computers, or microprocessors, if we wanted an example of a common, familiar machine. Most people today use a microprocessor to tell the time, not a mechanical clock. So it is remarkable that clocks still appear to be the

exemplar case of a machine in philosophical discussion of organisms. The reason for this is surely more genealogical than analogical; clocks are our intellectual inheritance, and not now particularly apt sources of (dis)analogies with organisms. Throughout this thesis we will see the claim that organisms should not be understood as machines justified by their difference with clocks.[1] I agree that organisms are not much like clocks, but deny that this fact has any particularly interesting implications about organisms and machines generally speaking. To properly understand the MOD today I will need to discuss the relationship between organisms and clocks, but to appropriately resolve the issue we will need to move past clockwork. The sorts of machines previously available for comparison to organisms, and on the basis of which historically significant MODs were justified, are now largely outmoded.

On the one hand, present technologies are far more diverse and sophisticated than the clockworks imagined by Descartes, Kant and Leibniz. Recently, Riskin's (2016) *Restless Clock* argued that we misunderstand Descartes because we fail to appreciate how much his (and his contemporaries') view of machines was "active" and "agential". However life-like the marionettes may have been to Descartes, our robotics are more so. Granted, what we find in biology we often do not find in technology, or find only in simplified imitation. Animals could fly and flash, count and calculate, balance and burrow, long before we could produce technologies that approximated these. But that is changing. Machines exist today that can fly, flash, count, calculate, balance and burrow. Indeed, some machines can speak, care for the elderly, derive theorems and self-replicate, more or less.[2] These new machines will form one side of the naturalistic examples used to counter MODs throughout the thesis. Given this Holocene explosion of technologies, it is increasingly dubious what philosophical work can be done by noting a difference between present machines and organisms.[3] An optimistic induction—which may nonetheless be apocalyptic—is that present differences between machines and organisms are also likely to be overturned, and so too would any philosophy founded on these.

On the other hand, discoveries within molecular biology and microbiology have shed greater light on just how diverse organisms themselves are. Biology is a science that collects counterexamples. The ancient divisions of macroscopic organisms into two kingdoms were just as soon amended to three (with the discovery of microbes), as they were expanded to four (with the distinction between prokaryotes and eukaryotes), then five (adding fungi). Now molecular phylogenetics has revealed hitherto unknown and still largely unexplored and unculturable diversity across a massively complex network of life. Some organisms appear

[1] For example, see Ch.5 Sec.1.

[2] See Ch.3 Sec.4.2.

[3] For example, Ch.5 argues that such differences do not speak conclusively in favour of, or against, the new mechanical philosophy.

to live autonomously, feeding at the bottom of food chains on raw materials, while others depend integrally on complex communities and cannot live alone. Some organisms require, or can accept, high temperatures, fast reactions and chaotic chemical processes, while others made do with simplified life-cycles, dividing slowly in glacial ice.[4] Some organisms seem to be able to replace any of their parts and live indefinitely provided the resources to do so, while others heal no wounds, live ephemerally or commit cell-suicide at the earliest signs of trouble.[5] Some survive by tooth and claw, others build elaborate structures and take up weaving.[6]

Most organisms are very different from us, most are unlike animals generally. O'Malley (2014) has pointed out that attempting to understand evolution writ large by (narrowly) focusing on our favourite macroscopic examples of life – ourselves, rabbits, butterflies etc. – neglects both the majority of life and its ancient origins, since both are microbial (O'Malley and Dupré 2017). Likewise, when attempting to understand the relationship between machines and *organisms broadly speaking*, focus on these "charismatic megafauna" neglects the majority of diversity in the latter. Some of these newly discovered and characterized forms of life form the other side of the naturalistic examples used as counters to the MOD throughout the thesis.

For example, autonomy is supposed to provide us with an understanding of organisms and an MOD (Ch.3). Considering animals, with their largely autonomous behaviours and comparing these to comparatively less autonomous machines – even to walking and talking robots – autonomy seems to mark a fundamental difference. But any theory of autonomy will need to contend with discoveries of diversity in autonomous capacities, organisms that simply cannot survive alone, perhaps because, like we ourselves, they cannot synthesize all the molecular components they require for survival. You will become ill and eventually die without a source of vitamin C; your cat will not, since it can make its own vitamin C on a diet of meat alone. Once such diversity in organisms is included in the comparison, autonomy fails as an MOD.

Nonetheless, this is not to say that all machine-talk in biology is appropriate. The rise of molecular biology and its reliance on machine-metaphors—DNA "coding", molecular "machines", the "programs" of development and gene regulation, the circadian "clock", molecular "motors" etc.—has lead to justifiable scepticism that a machine-based view of organisms is the right one. A torrent of philosophy of science has reacted to what it sees as an excessive or wrongheaded Cartesianism, which sees organisms as machines, and has

[4] Ch.4-5, see Mariscal, C., & Brunet, T. D. P. (2020). What Are Extremophiles?. Social and Conceptual Issues in Astrobiology, 157.

[5] Ch.4

[6] Ch.6

marshalled current biology against this equivocation.[7] Within the philosophy of biology at least, this current has been largely unopposed[8] and roundly supported by scientists themselves (Moore 2012, Woese 2004).

My aim in this dissertation is to provide a countercurrent, not as a defender of Descartes or machine-talk in contemporary molecular biology – which do not need defending anyway – but as a critic of the critics. To say that molecules or organisms should not be described as machines is one thing, but to say that this normative claim is justified on the basis of a fundamental MOD is another. This thesis collects together countless cases where an initially sensible claim – sometimes even a trivial one – that fits well with contemporary biology and engineering science, has been aggrandised into an MOD that does not. In these moments it may be little more than custodial work on the factory and laboratory floor, but it is necessary work if we are to avoid the philosophical pitfalls of neglecting contemporary diversity.

This dissertation is also a plea for naturalism about machines to go hand-in-hand with naturalism about organisms. Unfortunately, this is not always so, and to the detriment of our analysis of organisms. We are often good, at least well-intentioned naturalists about organisms, but fail to appreciate "late twentieth century machines" (Haraway's (1991) phrase) as newly discovered *natural* phenomena. My hope in noting this early on is that recognizing this can balance out whatever part of the arguments for and against MODs that seem to depend on issues of the appropriate philosophical methodology. Not all parties are innocent of errors; the primary error encountered and criticized in this dissertation is methodological. In attempting (with rigour, clear intellectual conscience and good-will) to provide a naturalistically grounded understanding of organisms, philosophers often rely on a concept of machines that is derived, not from naturalistic investigation into machines as phenomena, as parts of nature, but from pre-theoretic and pre-19^{th} century intuitions about machines. Though in different terms and in the philosophy of technology proper, Frederick Rapp makes this point of caution well.

> The first characteristic of technology is that it is always a factually given phenomenon. Its actual features cannot be deductively derived from a contemplation of its logical, timeless essence while disregarding concrete empirical evidence. To avoid arbitrary and nonconvincing speculations, philosophical analysis and reflection must be given based on contingent facts.
>
> —Friederich Rapp (1981)

[7] As we will see, the view that organisms are literally machines is neither particularly Cartesian (see Riskin 2016), nor sufficiently prominent to be pernicious within the science of biology. It is, however, a common feature of Western unscientific folk biology.

[8] cf. the conciliatory views of, e.g., Lewens (2013) or Hacking (1998)

Neither Rapp nor I note this for its profundity. I note it because contemplation of logical, timeless essence while disregarding concrete empirical evidence about machines leads to mistakes about organisms. Worse, the concept of machine that is supposed to help us understand organisms through critical argument is usually a very simple one. The bare-bones of Newtonian intuitions about how, e.g., *a clock* must work is a recurring theme in disanalogies with the physiology of multicellular organisms. One factually given phenomenon (organisms) are being understood by distinction with an intuitive conception (of machines). Indeed, sometimes arguments turn out not to hold up even to empiricism *about clocks*,[9] since their assumptions about how clocks must work do not match how clocks happen to work. If you want to argue naturalistically about how organisms should be understood (in some way or other) because they are different (in some way or other) from machines, then some prior understanding of those machines is necessary; advocates of an MOD also need to go out and observe machines. This is perhaps the more ironic point: it is because philosophers providing or relying on an MOD have aimed at a naturalistic analysis of organisms without having a naturalistic analysis of machines that they fail, overall, to have a naturalistic analysis of organisms. Rapp's caution has not been heeded.

The MOD encourages thoroughgoing naturalistic philosophers to temporarily neglect the staples of their philosophical methodology. Throughout this thesis I encounter cases where some traditional philosophical topic[10] rests on an MOD established on the basis of intuitions about a small sample of machines. In part, this is surely due to the seemingly excusable fact that this thesis addresses philosophies of *biology*, which are not in general also intended as philosophies of technology. If the consequences of the MOD were only a little unfairness or parochialism towards machines, this might be excusable. But this thesis shows that it is not so: the failure to treat machines naturalistically does continual disservice to our understanding of organisms.

1.3 Machines in the Philosophy of Organisms

I have explained what an MOD is, where they come from, and why we should care about them strictly for their role in our analysis of organisms. There are, nonetheless, independent reasons to be concerned with the MOD. There are two elephants in the room: (1) the emergence of

[9] See Ch.5, Fig.5.3.

[10] For example, the neo-Kantianism of Nicholson (2013) in Ch.2, and that of Maturana and Varela (1980) and Moreno and Mossio (2015) in Ch.3, the naturalistic metaphysics of Dupre and collaborators (Dupre and Nicholson 2018) in Ch.4 and the philosophy of mind of Godfrey-Smith (2016) in Ch.5. All are thoroughgoing naturalistic philosophers, yet all rely in part on characterizations of machines that are at best intuitive.

automata, organism-like, human-like, machines, and (2) the emergence of organism-machine hybrids, prostheses or cyborgs in the extreme.[11] These cases have fascinated the public and philosophers for centuries, and seem to present independent grounds for rejecting the MOD if only they exist. Reasoning about these cases is simple, powerful and has been influential: there are intermediate cases, such as organism-machines, machine-organisms, mechanical-persons, or organisms on the outside with machines on the inside, *therefore* there is no sharp difference between machines and organisms—organisms are not special, not fundamentally different from machines. Haraway perhaps put it best,

> Late twentieth century machines have made thoroughly ambiguous the difference between natural and artificial... and many other distinctions that apply to organisms and machines.
> —Haraway (1991) p.120

I agree with the conclusion and take the premises to be true, but see the reasoning as invalid. It is invalid because the MOD is prior to our analysis of these elephants in the room, *a fortiori*, since having some MOD in hand is required before we can make sense of claims about machine-organism intermediates.

Before embarking on a critique of a variety of different MODs (Ch.2-7), I want to discuss these points and bring them out into the open, showing how the MOD affects them rather than assuming it is the other way around. I also want to discuss them as issues arising firstly within our philosophy of organisms—for all their other charms, scientific plausibility and postmodern metaphorical uses—automata and cyborgs are a way of challenging our conception of organisms, ourselves in particular. The following sections discuss the connection of automata and cyborgs with the MOD (§ 1.3.1-2). Though these examples are more uncertain and hypothetical at present, they are continually connected with the MOD and will merit occasional discussion throughout the thesis. The final section of this introduction (§ 1.4) will cover in outline the content of each forthcoming chapter, what MODs appear therein, and which authors' views are addressed.

[11] There are more elephants, cases where the MOD is directly relevant, which I do not extensively discuss. For example, there is the issue of the ethics of treatment of machines, e.g. their capacity for organism-like mental states and thus merit for moral consideration; and there are the ethical and ontological issues associated with artificial personhood generally, perhaps most contentiously in the case of artificial sex-workers. Other interesting topics needed to be avoided. I set aside ethical issues, and issues of the definition of "life" generally speaking. I also set aside MODs that pertain specifically to the relationship between AI and brains. I find this focus on a single organ to the exclusion of the rest unhelpful to the broader aim of this thesis, however, my work on this topic has been published elsewhere (Brunet and Halina 2020).

1.3.1 1633, Automata and All That

It is not difficult to weave machines into a history of ideas, connecting the origins of modern philosophy, evolutionary biology, and contemporary catastrophic predictions about the human species—they all engaged with automata in some way. Quoting selectively, one can make Descartes, Darwin and Samuel Butler sound like the same person at different times. This section begins with these three threads of inquiry into automata, then shows how our grasp of them depends on MODs.

Descartes is a good place to mark an origin of early modern European philosophy. Descartes' philosophy deployed supportive analogies between organisms and machines. His mind-body dualism, and therefore his distinction between humans and other animals, rested on a conception of the body by analogy with machines like pulleys and waterworks, machines of his day, and on the assumption that God could construct machines with infinitely more sophistication than those available to him.

> We see clocks, artificial fountains, mills, and other similar machines which, even though they are only made by men, have the power to move of their own accord in various ways. And, as I am supposing that this machine [which imitates the movements of humans] is made by God, I think you will agree that it is capable of a greater variety of movements than I could possibly imagine in it, and that it exhibits a greater ingenuity than I could possibly ascribe to it.—Descartes (1633)

Switching from artificial beings to natural ones, Darwin employs remarkably similar reasoning to Descartes, however, Darwin's God was Nature. Darwin's evolution by natural selection rested on a conception of artificial selection of organisms like pigeons and orchids, topics current in his day, and on the idea that Nature, being far more powerful and of longer standing than man, was capable of much more sophisticated selection than those available to him.

> Slow though the process of selection may be, if feeble man can do much by his powers of artificial selection, I can see no limit to the amount of change, to the beauty and infinite complexity of the coadaptations between all organic beings...which may be effected in the long course of time by nature's power of selection.—Darwin (1859)

Both Darwin and Descartes make their case by extrapolating from familiar cases known to affect or change some complex entity (organisms or machines), by referring to a higher power or process that ought to be capable of much more than present in the familiar cases. Both

of these lines of reasoning are brought together, in unholy matrimony, in later predictions about the catastrophic effects of machines, for instance, in Samuel Butler's *Darwin Among the Machines*.

Butler evidently understood Descartes and Darwin both. For Butler, "subservience to the use of man has played that part among machines which natural selection has performed in the animal and vegetable kingdoms" [ibid]. Butler worries about the future of the human species, and applies a power similar to that of Darwin's Nature, not God, to the construction of sophisticated machines (Ch.6), to the "development of mechanical life," and ultimately to its dominance over us.

> Day by day, however, the machines are gaining ground upon us; day by day we are becoming more subservient to them; more men are daily bound down as slaves to tend them, more men are daily devoting the energies of their whole lives to the development of mechanical life. The upshot is simply a question of time, but that the time will come when the machines will hold the real supremacy over the world and its inhabitants is what no person of a truly philosophic mind can for a moment question.—Butler (1863)

On this last point, plenty a "truly philosophic mind" has questioned the idea that machines will develop so significantly, and these minds remain split as to whether it is a serious existential risk or an idle fantasy. Either way, we can see that the very same sorts of hypotheses, about the capacities of autonomous machines, appear in philosophy as early as Descartes and, as much as with Butler in 1863 as today. Moreover, automata are a topic of philosophy even when the apocalyptic ludditism is left out. Even if machines never overcome humanity or start a war, they are challenges to our conception of ourselves and have been since Descartes. Even if we are not worried for our lives, we remain concerned with our identities. Descartes' spiritual answer—that humans are different for their possession of souls—is no longer satisfying to a secular audience. Autonomous mechanical bodies are now possible, and they are possible on grounds more secure than the omnipotence of God or the near omnipotence of Nature. They are possible because they are more or less actual. Butler would be horrified.

Whatever way we lean on the possibility of fully-fledge machine automata, our analysis of them as a special sort of entity requires we make some distinctions between machines and organisms. It presumes some MOD. Surely, when Butler claims that "mechanical life" or "machines" will hold supremacy over the world, he does not intend a conception of machines that includes animal life. On this trivialization of Descartes, machines already hold supremacy over the world, so it is not a "question of time" before this happens. To

make sense of the claim as intended, we have to provide an MOD.[12] In Ch.5 I address Descartes' mechanical philosophy and subsequently developed, new-mechanistic, accounts of organisms, to show they likewise fail to provide an account of machines or an MOD.

To see how different MODs affect our analysis differently, we can consider another philosopher at the origin of the European tradition. Kant provided an MOD (Ch.2-3) that has received continued attention, accords well with common sense, and is prototypical of MODs generally. For Kant, machines are defined by the purposes they serve for organisms, while organisms have purposes, *ends*, only for themselves. At first this distinction seems to work very well. Hammers drive nails, clocks tell time, but organisms do what they do only to survive and reproduce their type. Birds are not *for* the songs we value from them, nor are beavers *for* making hats—they are *for themselves*, if anything. Moreover, this sort of distinction makes posing the kinds of questions we have about machines and organisms possible. *"Can machines hold supremacy over the world?"* becomes, on a Kantian reading: *"Can the things that are useful for us come to hold supremacy over the world?"* To which we might respond, in a wide reading of 'can', that it is conceptually possible for things that are useful for us now to later be in control of us.

Forget whether Kant is right and consider the role that the conception of machines in an MOD plays in interpreting Butler's philosophical question. If Kant's MOD is right, then Butler's idea seems to require that machines can become organisms provided they come to have ends in themselves. Or, at least, Kant's MOD entails that we can become the "machines of machines", provided we are reduced to our usefulness for them. Turning the idea that humans are "ends in themselves" on its head in this way also occurred to Nietzsche, shortly after Butler.

> It is with men as with charcoal-kilns in the forest... Mankind mercilessly employs every individual as material for heating its great machines: but what then is the purpose of the machines, if all individuals (that is to say mankind) are of no other use than as material for maintaining them? Machines that are ends in themselves – is that the *umana commedia* [human comedy]? —Nietzsche (1987) 585

Moreover, Butler's explanation of how this comes about by "subservience to the use of man" suggests that it is the very thing that makes machines what they are, "usefullness" on Kant's account, that puts them in the position to become organisms. Our concern about the potential existence of automata of the sort imagined by Butler must be founded on some MOD, and

[12]There are of course ways around having some analytically satisfying MOD and making similar claims. Someone wishing to advocate for a view like Butler's, might simply claim, for instance, that "entities made from gears and wheels will some day hold supremacy over the earth". However, substituting definitions for lists of current components is a very limited way of making sense of claims about machines.

if this MOD is like that provided by Kant, we are left asking whether machines can come to have certain sorts of purposes. This idea is addressed more fully in Ch.2. For now, note that a seemingly innocuous MOD, based in Kantian teleology, turns our line of inquiry away from future contingencies and towards our conception of what machines and organisms are at present.

1.3.2 1960, Cyborgs and All That

In 1960, Manfred Clynes and Nathan Kline wrote *Cyborgs and Space*, a speculative article for *Astronautics* detailing their view of how space-flight could be assisted by bodily modification of astronauts. Its influence – on science, philosophy and society – can hardly be doubted.[13] Many have found some challenge to the status quo in the cyborg concept. Clynes and Kline themselves evidently saw the possibility of modification of human beings as a vindication of the cybernetics[14] of the day; Donna Haraway's (1991) *Cyborg Manifesto*[15] used the concept as a metaphor for the intersectionality and changeability of race and gender identity; Ian Hacking (1998) saw a way to retroactively justify Canguilhem's (1947)[16] philosophy of machines and organisms, offered more than a decade earlier than Clynes and Kline (1960); Steve Mann and Hal Niedzviecki (2001) see the cyborgs of the present as changing human identity and potential; and Andy Clark and David Chalmers (1998) see cyborgs as changing, by extending, our minds—and all these presumably more so in a cyberpunk future. For the present connection, its importance lies in the fact that it spawned another current within philosophy that opposes MODs.

Cyborg concepts and actual cybernetic technologies are an inspiration for this thesis and will reappear occasionally therein. Cybernetics provides an interesting class of counterexamples[17] to some MODs, but it is not necessary for arguing against MODs. Microbiology also provides many interesting counterexamples,[18] and so does ornithology,[19] community ecology, robotics and computer science.[20] Cybernetic hearts challenge the way we think of autonomous human development and organs as human parts, but the existence of chemolithoautotrophic bacteria in specialized organs of deep-sea worms has the same effect,

[13] See § 6.3

[14] Surviving today as "systems theory".

[15] Which was still being offered for free, tucked inside alternative magazines, in the 1990's when the author obtained their first copy.

[16] See Fig.7.8

[17] Ch.2 § 2.4.1

[18] See Ch.6.

[19] Ch.7

[20] Ch.3

when considering (potentially non-human) organisms generally speaking.[21] As machine-organism chimeras, cyborgs are a lovely challenge for any MOD to prove its mettle. However, the MOD would fail even if we left cyborgs mostly out of the picture, focusing instead on the status quo of engineering and biology. The MODs under consideration here often fail simply because they fail to provide an analysis of organisms that holds up to discoveries within biology.[22]

Moreover, cyborgs are also not as powerful a counter-example to the MOD as its advocates sometimes suppose. The standard line that cyborgs threaten the distinction between machines and organisms can be turned around: if the MOD fails, then the very concept of the cyborg is threatened. The more difficult it becomes, according to our preferred theories, to establish an MOD, the more dubious any special category of "cyborg" becomes. Our futurism about cyborgs could just as well have gone another way, if we begin at the rejection of the MOD. Instead of the hypothesis that the future will hold more and more cyborgs of increasing sophistication—further "cybernetic" implants and "mechanical" appendages—we might imagine that the steady progression of discoveries in biology and engineering will turn the tables on our imagination. "Cyborgs" will increasingly become difficult to imagine, unclearly defined entities, somewhere in the middle-ground on a multi-factorial spectrum of diversity. This is so, I will argue, because organisms and machines are already in that position.[23]

Nonetheless, something interesting comes from bringing cyborgs in. In the final substantial chapter of this thesis I examine the idea that something common to the cyborg literature and, oddly, the literature on the evolutionary genetics of animal artefacts, furnishes us with a plausible relationship between machines and organisms. Both the iconoclastic advocates of cyborg ideas and the ultra-conservative accounts of organismal phenotypes have come to share the idea that animal artefacts can be seen as *extensions* of organisms.[24]

> Altering man's bodily functions to meet the requirements of extraterrestrial environments would be more logical than providing an earthly environment for him in space... Artifact-organism systems which would *extend* man's unconscious, self-regulatory controls are one possibility...
>
> The Cyborg deliberately incorporates exogenous components *extending* the self-regulatory control function of the organism in order to adapt it to new environments.
>
> —Clynes and Kline (1960), emphasis added

[21] § 3.4.1

[22] This excludes Ch.4-5, where I argue that *both* new mechanism and processualism provide an analysis of *both* machines and organisms, despite the contrary – MOD based – views of their defenders and critics.

[23] See the *Conclusion*, Ch.8.

[24] Ch.7

A bird extends its phenotype, its body, by building a nest, in much the same way that an hypothetical astronaut extends her phenotype, her body, by the building of a ship, or the implantation of devices for physiological control. Perhaps a robin is a cyborg, if nests are machines. Regardless, though this does not provide a *distinction* between machines and organisms as a whole, I argue (§ Ch.7) that the notion of extension does provide a *relationship* between machines and organisms that does not suffer the same pitfalls as the MODs.

1.4 Chapter Overview

We can now look at the content of the MODs later examined. The following list clusters the ideas addressed in this thesis into six kinds, corresponding to the chapters in which they appear and restated more simply for the time being, followed by the authors who most clearly advocate or rely upon them.

1. Teleology: *organisms enjoy a particular kind of teleology that machines do not; organisms have their teleology "intrinsically" while machines have their teleology "extrinsically"; organisms work or are of use only to themselves, machines work or are of use only to organisms; in organisms the parts are for one another, in machines the parts are for some external purpose*. This idea appears recently in Nicholson (2013). This is examined in Ch.2-3.

2. Autonomy: *organisms are defined by their autonomous organization; machines are non-autonomous because their organization differs significantly from that of organisms; moreover, the teleological differences between organisms and machines are a consequence of their fundamentally different organizations*. This idea originates in Kant's *Critique of the Power of Judgement*, from which it is adapted in Maturana and Varela's (1980) *Autopoiesis*, and more recently in Mossio and Moreno's (2015) *Biological Autonomy*. This is examined in Ch.3.

3. Processualism: *organisms are correctly understood as processes, whereas machines are best seen as mechanisms; organisms have a number of features as processes that machines lack, e.g., organisms are unstable but stabilized whereas machines are intrinsically stable, or organisms are subject to processes of turnover at all hierarchical levels, while machines are typically static structures*. This idea arguably has ancient origins (in the West) in Heraclitus, but has been defended by a number of authors in the 20th and 21st century (Rescher 1996; Seibt 1997; 2004; 2009, Dupré 2013; 2018, Dupré and Nicholson 2018). This is examined in Ch.4.

4. New Mechanism: *The mechanical philosophy incorrectly reduced organisms to machines or equated organisms with machines; organisms are fundamentally stochastic while mechanisms are not; organisms cannot properly be understood within the philosophy of mechanism, new or neo-mechanism, or the "mechanical philosophy", while machines can be understood therein; or, on the other hand, only new mechanism can correctly account for the multiple realizability of organisms, while a machine-based ontology cannot*. Mechanism arguably begins with the mechanical philosophy of Descartes, although receives particular attention as source or target of MODs at the rise and contention surrounding organicism and neo-mechanism (Needham 1928) and new-mechanism (Machamer et al. 2000; Bechtel and Abrahamson 2005; Dupré 2013; Skillings 2015; Godfrey-Smith 2016). This is examined in Ch.5.

5. Evolution: *Organisms evolve and machines do not; or, organisms evolve in some specific and rich sense while machines evolve only in a general and abstract (uninformative) sense.* The applicability of natural selection to change in technologies begins in Butler (1863), and the evolution of (human and non-human) artefacts generally has been contested in cultural evolution and its philosophy. I largely deal with the views of Lewontin (1970), Sterelny et al. (1996) and Lewens (2015). A different conception of evolution appears in Woese and Fox (1977) and Woese (1998), which I argue is applicable to technological evolution. This is examined in Ch.6.

6. Extension: *Machines as human animal-artefacts differ from non-human animal-artefacts in some fundamental way; machines are not extended phenotypes while non-human animal artefacts are.* This idea arguably begins as an MOD in Dawkins' (1982) *Extended Phenotype*, where he argues that some human artefacts, e.g. buildings, are not extended phenotypes of humans. On the other hand, before Dawkins, Camguilhem's (1952) notion of extended organs and Clynes and Kline's (1960) notion of the cyborg as exogenously extended provide earlier examples where extension was treated not as an MOD but as a relationship between machines and organisms. The idea has been taken up more recently by Haraway (1991), Hacking (1998), Clark and Chalmers (1998) and R. Kline (2009). I argue that, in our classification of biological things on the basis of part-whole relationships, we should adopt a broad view of what counts as an *extended part*, including both human artefacts, non-human animal artefacts, and social organisms. This is examined in Ch.7, the final substantial chapter.

Tactics differ between chapters but the aim is much the same in each: to argue against prominent incarnations of the MOD. The chapters cluster in pairs (2-3, 4-5, 6-7) based on their philosophical themes. In the first two substantial chapters (2-3), autonomy and

teleology have been offered explicitly as MODs, so my tactic is analytic refutation by way of naturalistic counter-examples. The teleological properties raised in Ch.2 are thought to be founded in an important way on deeper issues about the internal make-up of living things, specifically, their autonomous organization. So Ch.2 leads to the neo-Kantian accounts of organisms discussed in Ch.3.

In the next two, (Ch.4-5) process and mechanism-based ontologies have been offered as metaphysical theories, bearing directly on biology and entailing or deploying MODs as justification for their view of organisms. There my tactic is synthetic and conciliatory, showing that the use of MODs in both process ontology and neo-mechanism is misplaced. Neither requires nor benefits from the attempt to distinguish organisms from machines, and both are better off leaving the MOD out of their view. Since mechanism and processualism are currently seen as competing metaphysical accounts of biology, they are discussed sequentially.

In the final two chapters (Ch.6-7) I offer parallel positive theses about the inclusion of machines into biological theory; arguing that theories which include machines as objects of evolutionary and physiological study have not gone far enough to that end. Ch.6 argues that changes in machines can be genuinely evolutionary, according to three respectable accounts of evolutionary change. Moreover, that since a portion of the changes in machines are genuinely evolutionary, a complete understanding of machines requires an evolutionary understanding of them. Finally, and perhaps most speculatively, Ch.7 argues that machines can be included into biology as extended parts of organisms, generalizing the Dawkinsian theory of the extended phenotype to provide a foundation for claims that machines are extended parts of us.

Many have recently argued that we should not understand organisms by their relationship (or similarity) to machines (e.g. Nicholson 2013; Skillings 2015; Godfrey-Smith 2016). In arguing against these MODs, I agree with this position in broad strokes: we should not understand organisms by their relationship of *difference* to machines, since our best accounts of these differences turn out to fail and thus furnish little understanding. Given machines and organisms are plainly similar in a number of respect and different in others. What I argue is not that organisms *are* machines, or that old machine-based analogies about organisms provide special understanding of them, but that none of the proffered differences amounts to a fundamental distinction. This is reviewed in the concluding Ch.8.

In this introduction I hope to have enticed the reader in two ways: the MOD is pervasive and, whether we are interested more in providing an account of organisms or more in speculative futurism, the MOD has serious consequences. The next chapter begins a two-chapter treatment of MODs based on an organizational account of teleology.

Chapter 2

Teleology and The Machine Organism Distinction

Abstract: In this chapter I argue against the view offered by Nicholson (2013) that the difference between intrinsic and extrinsic forms of teleology is the MOD. According to Nicholson, organisms are characterized by having intrinsic teleology—their ends, telos, or purposes are directed towards themselves—while machines are only extrinsically teleological. This is an overly idealized picture of the MOD. There are machines that operate towards ends or purposes that benefit nothing but that machine and there are organisms that (through external control or symbiotic evolution) have come to have purposes directed towards others.

2.1 Introduction

Many hold the view that machines are defined by their relations to external, human intentions, while organisms have purposes or teleological properties in some rich internal sense. On this view machines only have a teleology by proxy, through their relationships to us. Nicholson (2013) offers the best representation of this view, arguing that it is the difference between the 'intrinsic teleology' of organisms and the 'extrinsic teleology' of machines that fundamentally distinguishes them. I here argue that this binary division of machines and organisms into intrinsically and extrinsically teleological is incorrect. The teleological profile of entities should instead be seen as a spectrum (Fig.2.1).

In section 2 I reiterate Lewens's (2013) argument that on an etiological account of what it takes to be an artifact or organism there are many examples of entities that are both—ourselves included. In section 3 I offer an exposition of the alternative view, presented by Nicholson (2013), that it is not etiological, but teleological differences that distinguish

machines and organisms. Sections 4-5 argue against this view: first through examples of extrinsically teleological organisms and second through intrinsically teleological machines. In the last section (6) I present a view of both biological and technological things where entities can have a mixture of both intrinsic and extrinsic purposes.

2.2 Etiology

The best account of teleology is etiology. This is a contentious claim; one with a history too long to address here.[1] Philosophers of biology predominantly accept that theological explanations of teleology are wrong and that the correct account of biological teleology must somehow involve reference to evolutionary history, but disagree on precisely how (Godfrey-Smith 1993). For my purposes, De Sousa's (2017) characterization of the status quo will suffice,

> [T]he concept of objective teleology—independent of human interests and purposes—does not require intelligent design after all. Natural functions can be identified with those effects of an organ's activity that resulted in its being selected for, and hence explain its present existence... the aetiological explication of natural function marks one of the few genuine advances in philosophy in the past hundred years. —De Sousa (2017) p.147

This etiological or Selected Effect (SE) account of function and teleology has not gone without its critics and is not without counter-examples of non-selected yet intuitively functional things. Nonetheless, to my mind it is our best account of biological teleology. I will here consider an etiological account of teleology in machines and organisms first, before arguing against an alternative to it, offered by Nicholson (2013) and founded in part on the organization account of functions (see Moreno and Mossio 2015; Ch.3).

An etiological account can also be given for the functions, teleology, of machines or artifacts—even when human interests and purposes are involved (Lewens 2006). We can always ask, of an artifact or its parts: *What past effect of this thing explains its present existence?* When ordinary artifacts are in question, it may not be the same sort of explanation

[1] However, it is worth noting that the idea that evolutionary etiology implies the absence of purposes in nature goes back much farther than is typically said in quick summaries of the history. It is often said that the etiological approach dates back to Wright (1973) or Millikan (1984) but, surprisingly, Nietzsche (1881) in Daybreak already sees evolutionary explanations as an alternative to natural purposes, "The impartial investigator who pursues the history of the eye and form it has assumed among the lowest creatures, who demonstrates the whole step-by-step evolution of the eye, must arrive at the great conclusion that vision was not the intention behind the creation of the eye... and 'purposes' fall away like scales from the eyes!" (Book.II.122).

we obtain from natural selection, but what Wright (1973) calls "consequence-selection", a form of conscious choice or deliberation about intended effects. For example, a knife had the effect of cutting when pressed against certain surfaces, it is the intention of some humans to select tools that have the effect or consequence of cutting, so the effect of cutting explains the present existence of some knives. Thus cutting is a function, or the purpose, of some knives in an etiological sense based on consequence selection.

This does not entail that the intended effect is the same as the actual effect that in fact explains the present existence of a thing—our intentions can mislead as to why a given artefact exists. Weapons of self-defence are intended to have the effect of bringing about physical safety, but often do not have that effect, or have precisely the opposite effect. Here the actual past effect that explains the present existence of the weapon is perhaps that of producing the belief that it will bring about physical safety, for example. In section 5 I will discuss ascribing purposes to unintended effects in machines. For now, note that the separation of intended and actual effects does not change the overall form of an etiological explanation deploying consequence-selection. As Wright (1973) notes,

> Both natural and conscious functions are functions by virtue of their being the reason the thing with the function "is there"... The differentiating feature is merely the sort of reason appropriate in either case: specifically, whether a conscious agent was involved or no... it can be looked upon as a matter of mere etiological detail, nothing in the essential form of the explanation. –[ibid]

It is not the mere availability of an etiological account of teleology that can distinguish machines and organisms but, if anything, a distinction between sorts of teleological or etiological accounts. One way of distinguishing between sorts of etiology is by what sort of process is referred to in the explanation for present existence. Some explanations refer to processes of natural selection, some to processes of reasoning or intention. Let us call these *natural etiologies* and *rational etiologies* respectively. Rational processes are "natural" in the sense that everything is, but the distinction is helpful here. Now, we might distinguish machines from organisms as follows: if an entity is an organism then its functions are explained by a natural etiology, and otherwise for machines.

This distinction fails for many machines and organisms. It might work well enough to distinguish medium-size-dry-good machines and organisms, but there are others. There are machines and organisms for which an historical explanation of the processes resulting in their current existence do not neatly separate into the natural and rational. We can look to some of the more current examples from the special sciences of biology and technology. A genetically modified mouse is an entity that has both a natural etiology (explaining its evolution qua mouse) and a rational etiology (explaining its transformation qua experimental

research animal). Likewise, an evolved microchip—the product of an artificial selection of randomly generated microchips (see Thompson 2002; Lewens 2013)—has both a natural etiology (explaining its evolution qua chip lineage) and a rational etiology (explaining its copying and construction qua experimental machine). In such cases one and the same part may owe its existence to both a history of intention and natural history.

Furthermore, we needn't rely on 21st century examples. If we see artefacts as the products of rational etiology, then we already eroded that distinction prior to the industrial revolution. As Lewens (2013) notes about worries that synthetic biology "threatens to blur the distinction between organisms and artefacts",

> A dairy cow is an organism if anything is. And yet, dairy cows have clearly been modified by human breeders with the purpose of increased milk yield in mind. Dairy cows are organisms that have been purposefully manipulated: they are organisms and artefacts at the same time. Artificial organisms have been around for as long as intentional agents have practiced artificial selection.—Lewens (2013) p.642

Cows had evolved by natural selection without rational intervention for some time before humans began to breed them. They are organisms. But dairy cows have been artificially selected, humans have reasoned about which pairs to intentionally breed to achieve, for instance, maximum milk production or longevity. So they are artifacts. Likewise, we can look to domesticated plants, artificially selected and engineered to suit our desires, as Pollan (2001) does with potatoes, marijuana, tulips and apples. All of these organisms are artefacts; each has a partly rational etiology owing to artificial consequence selection.

We are also partially artificial in this sense. Humans perform a sort of "artificial selection" on themselves when intention and reasoning about the production of the next generation is involved. Cases vary from the morally reprehensible to the everyday—genocide, infanticide, forced sterilization, eugenics, trait-specific mate choices, sexual preferences and any case where reasoning about the biological characteristics of kin precedes mating. If sexual selection in humans involves any reasoning or intergenerational intentions whatever, then we are artefacts insofar as we have sexually selected traits. We are products of our own intentions and therefore a perpetually constructed *causa sui*. Although the distinction between natural and rational etiology is defined in part in terms of human intentions, while sexual selection is not, in the human case sexual selection implies rational etiology. On the view that the natural-rational distinction is the organism-artefact distinction, we are artefacts.

If there is a function or purpose based distinction between machines and organisms—between artifacts and us—it is not captured by the applicability of the distinction between

rational and natural etiology. But recently there has been a revival of interest into non-etiological accounts of teleology (see Walsh 2015; reviewed in Dupré 2017; see also Mossio et al. 2009; Ch.3), with the added hope that the distinction between machines and organisms can be better cast in another way (Nicholson 2013). The remainder of this chapter describes and argues against the view, offered in Nicholson (2013), that the MOD can be defined by intrinsic and extrinsic teleology.

2.3 Intrinsic and Extrinsic Teleology

Instead of rational and natural *etiology*, Nicholson (2013) offers a view of the distinction between machines and organisms that depends on the difference between the sorts of *teleology* applicable to each. My aim in the remainder of this chapter is to explain the teleological account and show how it likewise fails as an MOD. Like Lewens (2013), Nicholson (2013) is also concerned with the implications of synthetic biology for the distinction between organisms and machines,

> . . . If synthetic biologists eventually succeeded in engineering from scratch a living system—that is, a system that was self-organizing, self-producing (upon its initial creation), self-maintaining, and self-regenerating—then such a system, despite its artificial origin, would still have the capacity, by virtue of its internal organizational dynamics, to act on its own behalf in accordance with its own norms.—Nicholson (2013) p.674

He takes this as a *reductio* of the idea that machines are distinguished by an "artificial origin [etiology]". If a totally artificially originating synthetic living system were to be made—as in the above definition—it would be an organism on his account, not a machine. He instead offers a view based on varieties of teleology, which he believe is "far better suited than the distinction between natural and artificial origin" ([ibid], p.674). The essence and perhaps the best representation of this view is as follows.

> Organisms are intrinsically purposive, whereas machines are extrinsically purposive. A machine is extrinsically purposive in the sense that it operates towards an end that is external to itself. Its *telos* is imposed from the outside and it is of use or value to an agent other than itself. A machine does not serve its own interests but those of its maker or user. On the other hand, an organism is intrinsically purposive in the sense that it acts on its own behalf, towards its own ends. Its telos is internal, arising from within, and ultimately serves no purpose other than maintaining its own organization. –Nicholson (2013) p.671 his emphasis

This is the representation of Nicholson's view that I will be addressing. I take it to consist essentially in the following claims: (1) there is a distinction between sorts of purposes, intrinsic and extrinsic, (2) extrinsic teleology involves a purpose that is somehow directed externally, either by "operating towards an end external to itself" or by being "of use or value" to another agent, (3) intrinsic teleology involves purpose that is somehow directed internally, either by an agent "acting on its own behalf" or "towards its own ends", and (4) that the difference between machines and organisms is that organisms are intrinsically purposive while machines are extrinsically purposive. Only the last point is relevant to the issue of the MOD, so I will not dispute (1)-(3) while arguing against (4).

The distinction between intrinsic and extrinsic teleology also manifests, according to Nicholson, in two further interesting respects: through the attribution of functions, and the possession of certain self-directed capacities (e.g. self-organization, self-formation, self-production, etc., see Ch.3). These will be addressed first before turning to the more strictly teleological aspects of the view.

2.3.1 Intrinsic Teleology and Functional Ascription

Mindful of domestication and human "interferences" with other organisms, Nicholson (2013) notes that the difference between intrinsic and extrinsic teleology also gives rise to a difference in how organisms and machines are ascribed functions. Organisms, he says, do not have functions as wholes, while machines do. He explains,

> An organism does not have a function because its operation is not good for anything; it simply acts to ensure its continued existence. [In Footnote] This is true *ceteris paribus*. Humans do domesticate animals and cultivate plants, and in doing so use them for their own ends. Such human 'interferences' confer upon the manipulated organism a level of functionality it would not otherwise have in nature. –[ibid p.671]

However, what is true *ceteris paribus* is false *simpliciter*. The world that this *ceteris paribus* condition asks us to imagine is far from our actual world, and was so even before the evolution of human intention. We are asked to consider functional ascription to organisms in nature, where 'interferences', manipulation, domestication and cultivation are the sorts of interventions that make the functional profile of an organism unnatural, and these must be excluded to validate the claim about functional ascription to whole organisms. But if we are told that "organisms do not have function ceteris paribus" means "organism do not have functions unless there is something unnatural, rational, about their history", then—as well

as conceding to the etiological account—we can only say that some organisms do not have functions.[2] Plenty of others do.

Moreover, it is not only human 'interferences' that impart functionality to whole organisms, non-human cases abound where one organism has learned or evolved to confer a level of functionality on another. Consider Voltaire on "equality",

> What does a dog owe to a dog, and a horse to a horse? Nothing, no animal depends on his like; but man having received the ray of divinity called reason, what is the result? Slavery throughout almost the whole world. —Voltaire (1764)

Voltaire was not *overstating* the consequence of reason. Countless humans have a level of functionality conferred upon them that they would not otherwise have—they have their telos imposed upon them. Humans also impose functions on non-human animals. As Milan Kundera[3] put it, "'Man the cow parasite' is probably how non-man defines man in his zoology textbooks". We have been cultivators and domesticators since prehistory, but we are not the only animals that cultivate and domesticate. However, Voltaire was *understating* the teleological relationships between non-human organisms. Even prior to human history there were lineages capable of manipulating and intervening on others. A pack of dogs or a herd of horses must cooperate, they do depend on and make use of one another. Likewise, farming behaviors such as the ant-aphid or ant-fungus symbiosis result in relations of subservience between non-human organisms—slavery throughout an even larger proportion of the world. Organisms have long adapted to make others perform functions for them and this is a sort of biological relationship we should not neglect or exclude in service of an MOD. The consequences of this for the teleological view of the MOD will be taken up fully in § 2.4-5 below.

2.3.2 Intrinsic Teleology and Self-Organization

That organisms are intrinsically teleological should be explanans or explanandum for something that makes them different from machines. For Nicholson (2013), the explanation of intrinsic teleology in organisms derives from features of their organization. The main treatment of this topic is provided in the following chapter (Ch.3), though some background here is helpful for understanding the basis of renewed interest in an "intrinsic" account of organismal teleology as an MOD.

[2] These are precisely the same sorts of interventions that would grant something a rational etiology in the above sense, an etiology that does not support a machine-organism distinction. It is hard to see how Nicholson would need to make this admission on a strictly organizational account of function (see Ch.3), since it seems to deploy an etiological account to ascribe functions to organisms that have been intervened upon.

[3] Kundera, The Unbearable Lightness of Being (Faber and Faber ltd.) p.279

> Organisms, unlike machines, are not only organized but are also self-organizing and self-regenerating systems. Organisms are intrinsically purposive because they have an autonomous self: the phenomena of self-formation, self-preservation, self-reproduction, and self-restitution are all characteristic of the internal organizational dynamics of living systems.—Nicholson (2013) p.671

On this view, a hypothetical synthetic living system is still an organism because its internal organization explains its capacity to have an intrinsic teleology, despite its artificial origin or what use is made of it. I think this is the wrong way to establish the explanatory purchase of the intrinsically teleological account of organisms.

In Nicholson's distinction between intrinsic and extrinsic teleology above (§ 2.3.0), there is an implicit distinction between the *source* and *content* of teleological properties. The source of an organism's telos is internal, it arises from within, whereas the source of a machine's telos is external. The content of these ascriptions of purpose, however, relates to how they are "directed", on whose "behalf" they operate, and where their "use or value" lies. This presents a problem, because the source and content of purposes are, presumably, independent in some cases—something might have an internal source of purpose with external content—yet Nicholson presents the two together, as if both source and content made the purposes of organisms intrinsic.

How an organism can have an intrinsic source of teleology, purpose, is supposed to be related to its autonomous self-organization. This view of the source of teleology has antecedents in Maturana and Varela (1972) and more recently in Mossio et al. (2009) and Mossio and Moreno (2010; 2015), but can be dated back to ideas of reciprocal causality in Kant's *Third Critique*.

> [P]urposiveness is grounded either on the internal possibility of the object, or on the *relative* possibility of its external consequences. In the first case the teleological judgement considers the **perfection** of a thing in accordance with an end that lies in it itself (since the manifold elements in it are related to each other reciprocally as end and means); in the second the teleological judgement about a natural object concerns only its **usefulness**, namely its correspondence with an end that lies in other things. —Kant (2001, p.49), emphasis in original

In the following chapter (Ch.3) I will argue that autonomous self-organization (or self-directed capacities[4] of organization) does not amount to an MOD. This is in part because each self-directed capacity comes in degrees, and many are possessed to some degree or other by machines. Likewise, many organisms lack some self-directed capacities completely.

[4] E.g. self-organization, self-formation, self-regeneration, self-reproduction, etc.

Thus, if it is the nature of self-directed capacities such as self-organization that is the source of the teleological characteristics of entities, then some organisms lack features of intrinsic teleology and some machines possess them. Kant was unaware of the extent of interdependence of life, and did not live to see our present state of technological development. If he did, one would hope we would partially rescind his distinction between internal and relative purposiveness—organisms are imperfectly organized, and some machines are more than merely useful.

For Nicholson (2013), these organizational claims are "related in one way or another to this key distinction [between intrinsic and extrinsic teleology]", and this is a conclusion I am happy to affirm. However, if organization is a necessary condition for teleology, then the intrinsic-extrinsic distinction is not a special MOD, over and above that of self-organization. If the teleological MOD is to occupy a special place, then it must be related to organization by some way other than being a necessary condition—otherwise it fails as an MOD if organization does. This would be a setback, since Nicholson's view seems to be especially lodged in teleology. Instead of looking to special sorts of intrinsic organization, I here consider whether the division of machines and organisms into extrinsically and intrinsically teleological provides an explanation of the *content* of purposive claims, their ends, telos, "direction" or functions.[5]

Restricting attention on the teleological MOD to what I claimed above is its best representation, I will argue that some organisms have functions that are extrinsically teleological (§ 2.4) and some machines have intrinsically teleological parts (§ 2.5). Beyond the specific debate about the correct account of the distinction between machines and organisms, this allows us to address a pair of wider issues: whether or not organic function arises strictly from the intrinsic nature of organisms; and whether machine functions can be conceived as the mere product of, or constituted by, human intentions and purposes. I will answer both cases negatively and conclude with an alternative view.

2.4 Extrinsically Teleological Organisms

> The ubiquity of symbiosis and cooperation...affect the way in which researchers think about biology's most taken-for-granted entity, the organism. Microbial

[5] Another confusion here seems to arise from confounding properties that are intrinsic (i.e. non-relational) with internal (i.e. physically inside). Properties that obtain inside an organism can nonetheless be relational, extrinsic. Homeostasis, a patently organizational property, often depends on relations to external states of affairs, such as external physical or chemical variables. Self-directed capacities, being often physically internal properties, seem like a good candidate to explain the intrinsic properties of an organism, but need not align to the distinction between teleological properties and relations.

> ecology shows how collaborative interactions at many scales blur the usual distinctions that are made between so-called individual organisms and the larger organismal groupings of which they are a part.—O'Malley (2014) p.156

I will argue that some symbiotic relationships involve extrinsic teleology—functions of organisms for others. While O'Malley's (2014) aim is to show that a narrow "Dawkinsian" account of the organism is violated in micro-ecology, mine is to show that symbiotic ecological interactions generally speak against an intrinsically teleological account of organismality. Once we see how organisms live together, we see that they do not always act only on their own behalf or only towards their own ends. Nicholson (2013) recognizes but dismisses this fact in a footnote discussing the apparently extrinsic nature of the functions in *biofilms* and *colonial insects*. He writes:

> It is interesting to observe that in such cases one always encounters the added difficulty of deciding whether the systems in question constitute populations or individuals. Is a biofilm, or an insect colony, a community or a (super)organism? Regardless of the answer, the fact that the 'intrinsic vs. extrinsic purposiveness' distinction is strongly correlated with the 'individual vs. population' distinction provides a compelling reason for adopting intrinsic purposiveness as a means of individuating organisms. –Nicholson (2013) fn.4

The worry is that, in (some) biofilms and colonial insects, there are many organisms engaged in symbiotic relationships where some putative functions are *for* other members, and this seems to grant extrinsic teleology to some of the organisms living, e.g. in biofilms and colonies. However, both of these symbiotic systems are contested cases of individuals. Ereshefsky and Pedroso (2013) argue for the individuality of biofilms, and Wilson (1980) makes similar claims about colonial insects. Thus, we have some license to treat these larger-than-single organism systems as individuals, making the extrinsic teleology between colony members intrinsic to the colony as a whole individual. Given these biological outliers, perhaps there is no perfect correspondence between organisms and intrinsic teleology, however, Nicholson claims that there remains a sufficiently high degree of correlation to justify using intrinsic teleology as a means of individuating organisms.

Nonetheless, there are two problems with this approach to resolving the apparent extrinsically teleological features of symbiosis: (§ 2.4.1) there is a switch from the notion of organism to biological individual, and (§ 2.4.2) a focus on too narrow a class of symbioses (on only those that are plausible individuals). The following two sections argue that intrinsic purposiveness does not strongly correlate with biological individuality, nor does it provide a means of individuating organisms.

2.4.1 Intrinsic Teleology and Biological Individuality

Even if intrinsic teleology correlates with biological individuality, this does not imply that it correlates with organismality. This is because there are a variety of putative biological individuals at organizational levels above and below that of the organism. Genes, chromosomes, colonies, populations, species and clades, have all been argued to enjoy varieties of "biological individuality" (in a broad sense) and to be bearers of self-serving functions specific to their level of organization. None of these individuals are organisms.

Dawkins (1976) famously advocated an intrinsically teleological view of genes—selfishness in one sense—while Doolittle and Sapienza (1980) and Orgel and Crick (1980) were arguing that the evolution of transposable element genes did not require an organism-level explanation—selfishness in another sense. At some level above organisms, Hull (1986) has argued that species are best conceived as biological individuals—rather than classes thereof—and Jablonski (2008) has convincingly shown how selection and functional attribution can there apply. We need not accept all of these views of individuality or function, but unless we reject them all we cannot hold to the perfect correlation of organismality, biological individuality, and intrinsic teleology.

Perhaps we could save something of the intrinsic teleology view by admitting that the focus on organisms was hasty. Instead, perhaps intrinsic teleology is correlated best with "biological individuality generally speaking". All that seems to follow from a high degree of correlation between the 'intrinsic vs. extrinsic purposiveness' distinction and the 'individual vs population' distinction is that there is a compelling reason to adopt intrinsic purposiveness as a means of individuating biological individuals (and a reason that intrinsic purposiveness should not be adopted as a means of individuating populations of individuals).[6] Then, Nicholson's (2013) central claim could be modified as a denial that machines are biological individuals since, supposedly, only the latter are characteristically intrinsically teleological. However, this is a substantially weaker view than it might seem. For one, it still allows for machines to be biological. A number of biological things are not (or marginal, questionable) biological individuals (e.g. ecosystems, organs, and holobionts). On the other hand, as will be addressed in the following section, if we pick out individuals that are characterized by intrinsic teleology we find that some of them are not entirely biological.

[6] In Ch.7 I argue that reciprocal causation should be involved in an account of machines as extended parts of larger biological wholes. However, this is not an MOD.

2.4.2 Intrinsic Teleology and Individuation

Could we nonetheless use intrinsic teleology as a means of individuating organisms? Lets first consider how we might individuate biological parts generally. McShea and Venit (2000) and Newman (2006) offer a helpful view of the individuation of biological parts. In a large collection—network, web, graph—of connections between entities, we can pick out parts—sub-networks, sub-graphs—on the basis of which entities have more internal than external connections, provided we choose a way to establish connections. The entities that form a part will mostly be connected to each other, though they may have connections to other entities or other parts. When the part in question is an organelle we might take connections to be established by protein-protein interaction networks; when the part is a cell or tissue we might consider connections in terms of chemical or signal exchanges. In each of these biological cases, how we individuate on the basis of networks depends on the connectivity between parts. In the present case, we are concerned with teleological-individuals, so the relevant form of connection should be established by the relation between an entity and its ends—where a connection is internal to an entity when its "telos arises from within" and external when its "telos is imposed upon it"—so that connections match the ascription of types of teleology. I think we gain some traction here by looking at how teleology might serve as a means of individuating biological parts, and whether the parts individuated by intrinsic teleology turn out to be organisms.

Often, if we restrict ourselves to individuals of the organismal and super-organismal level, we can expect that such individuals have more intrinsic than extrinsic teleology, but this does not imply that anything with more intrinsic than extrinsic teleology will be an individual organism, which would be required to use it as a criterion of individuation for organisms. Nicholson (2013) is right that these teleological connections can carve out individuals, but wrong that these are organisms. Moreover, some such individuals are not strictly biological. Sometimes, connected collections of machines and organisms will have an intrinsic teleology.

Consider the view of machines on offer: they have an extrinsic teleology and function for things (organisms) external to themselves. Now, consider some organism together with the collection of machines that have functions for and only for that organism—perhaps a patient on dialysis or yourself and only those machines specifically for your ends.[7] Calling this collection an "individual" requires a somewhat promiscuous individuation, we might think. No matter, we are not concerned with a broad notion of individuality, but only the sort of individuality that arises due to a predominance of intrinsic teleology. Importantly for us, this collection has more intrinsic than extrinsic teleology: perhaps most of your bodily functions are intrinsic to you and the functions that your machines have for you are intrinsic

[7] For discussion of machine–organism amalgamations, see Ch.7.

to the collection containing both you and them. So a cyborg-like collection of an organism and some machines can sometimes be an intrinsically teleological individual.[8]

Here, the attempt to individuate on the basis of teleology results in the conclusion that many of us are a sort of cyborg—with all the attendant post-modern appeal. However, it does not provide a means of individuating organisms. Many maximally intrinsically teleological individuals turn out to be non-organisms—indeed, they are not even members of the usual hierarchy of biological individuals since they include machines as parts (see Ch.7). If we go looking for individuals that are characterized by intrinsic teleology, we find that many are not organisms. Next I will show that if we examine some organisms we find that they are not fully characterized by intrinsic teleology.

2.4.3 What is the function of hairs on pollinators?

Many organisms engage in symbiotic interactions that involve extrinsic teleology—functions for others. Tightly knit symbiosis can sometimes be good candidates for super-organismal individuals, but not all symbioses are so well integrated.[9] For instance, colonial insects engage in mutualisms with flowering plants.

> Bee pollinated flowers have evolved in such a way that a visiting bee has to brush against the flower's anthers bearing pollen... Compared with other insects, bees are extremely hairy. Each hair has a branched structure that makes it highly effective at catching pollen. —FAO[10]

Why are bees so hairy?[11] The best explanation of bee hairiness involves a coevolved symbiosis between bees and flowers where both have functions for which the other is the beneficiary. Explanations that do not involve this sort of extrinsically teleological relationship are available, yet they either ignore the reciprocally beneficial nature of the bee-flower relationship or ignore coevolution.

Here is an explanation that does not require extrinsic teleology in bees: *the function of bee hairiness is to allow them to capture pollen grains from the flowers they pillage, the benefit*

[8]Moreover, these collective cyborgs can demonstrate some of the self-directed capacities thought characteristic of autonomous organisms. There are some machines, such as dialysis machines, without which some organisms cannot perpetuate or regenerate, so only the collection containing both a patient with kidney failure and a dialysis machine could be self-perpetuating or self-regenerating. More in Ch.3.

[9]There are symbioses between hymenoptera and angiosperms that do seem to constitute instances of evolutionary individuality, such as the fig-wasp symbiosis, yet it should suffice to say that not all pollinator systems are like this.

[10]http://www.fao.org/tempref/docrep/fao/012/i0842e/i0842e04.pdf (July 25 2018)

[11]Perhaps it is better to ask, "Why do hives of bees have so many hairy members?" to account for the fact that individual bees often have no non-kin-selective reproductive success whatever and are not the beneficiaries of their own hairiness. This is not being overlooked, but simply ignored for stylistic reasons.

to bees is that this pollen provides them with food for their hive. The hairiness of bees does have the effect of carrying pollen from flower to flower, facilitating the sexual reproduction of angiosperms, but this is not a purpose of bees. This explanation is unattractive, since it ignores the importance of the reciprocally beneficial relationship that pollination has for bee colonies. It clearly is beneficial for bees that their primary food-source is better able to engage in its primary form of reproduction: the following season's harvest depends on this.

Accounting for the reciprocally beneficial effect of pollination on bees, we can provide another explanation of bee hairiness: *the function of bee hairiness is twofold, (1) it provides a physiological mechanism of pollen capture, (2) it maintains or improves the health of the environmental food source.* Here the functions of hairiness are still of and for bees. Black-boxing flowers, we obtain as explanation that locates all functions and the beneficiaries of hairiness in bees. While the initial explanation was unsatisfactory for ignoring the reciprocally beneficial effects of pollination, this explanation ignores the fact that flowers are more than a passive, constructible, environmental resources—it ignores the fact that the bee-flower symbiosis is coevolved. Indeed, it is highly idealized to treat the population dynamics of angiosperms as a "complex environmental effect" of the hairs of insects. It is tempting, at first, to give an organism-centric explanation for the purposes of structures in organisms, but this story is partial.

Accounting for coevolution of the bee-flower symbiosis puts us in a position to ascribe teleology to bees in the same way that we might ascribe it to paradigmatic machines. *The function of bee hairiness is still twofold, (1) it provides a mechanism of pollen harvest, and (2) it also has the function of pollinating flowers.* This is an effect that "operates towards an end that is external to itself", the telos of which is "imposed from the outside and is of use or value" to flowers. Flowers, for their part of the coevolutionary bargain, have their own extrinsically teleological function: providing sufficient pollen and nectar. This situation, wherein one organism bears a function benefiting another, is maintained by a reciprocal exchange of purposes. Indeed, it is not entirely a case of *self*-maintenance, *self*-production, or *self*-reproduction (Ch.3).

The effect of pollinating flowers is of course also valuable to bees, due to the coevolved mutualism. The Kantian objection here would be that the effects of co-evolved traits ultimately come back around to benefit the organism, so what seems like an extrinsic function is intrinsic after all—regardless of the fact that this effect propagates by a circuitous route external to the organism—it is reciprocal causation nonetheless. If we consider the ultimate effects of bee hairiness, the Kantian would note, we would see that these are ultimately for bees and not for flowers. This reaffirms the teleological view, with a qualification that

organisms "*ultimately* act on their own behalf" and "*ultimately* towards their own ends". However, the Kantian objection is wrong on two fronts: one biological, the other conceptual.

Firstly, there is nothing preventing one partner in a mutualism from withdrawing its reciprocity and converting a former mutualism into commensalism or parasitism. Then, a structure from the other partner would still be valuable to the first yet not, for that reason, ultimately beneficial to itself. A flower, for instance, might begin to produce nectar and pollen which has no nutritious value to bees yet remain the beneficiary of bee hairiness.[12] As research on Hummingbird flowers suggests, when bee hairiness is not doing the trick, flowers evolve to cut bees out of the former mutualism through "anti-bee" and "pro-bird" morphological changes (Castellanos et al. 2004). Just as the *telos* of bee hairiness can be imposed from the outside, it can be deprived. Such cases show that organisms do not always ultimately act only on their own behalf, or ultimately towards their own ends.

Secondly, we need not see intrinsic and extrinsic purposes as mutually exclusive. Even when a trait is ultimately beneficial for the organism bearing it, this need not be overriding of any intermediate extrinsically beneficial effects. Even when symbiotic interactions ultimately benefit an organism, it is nonetheless true that intermediate effects are of use or value to others and that such interactions serve purposes other than maintaining their own organization. Pollinating flowers is an effect of bee hairs and it is valuable to flowers, all of this withstanding the cases when pollination is also valuable to bees. A given structure can have a diffuse spectrum of purposes. I return to this point below, but for now turn to the other case at hand: purposes in machines.

2.5 Intrinsically Teleological Machines

Above we saw that picking out intrinsically teleological individuals does not necessarily pick out organisms—sometimes we are left having individuated a cyborg. We also saw the inverse, that picking out an organism does not necessarily give an intrinsically teleological individual—symbiotic organisms can have extrinsic purposes. This establishes that intrinsic teleology and organismality are independent, but not that intrinsic teleology fails as an MOD. For that, we need intrinsic teleology in machines.

Below I argue that the design, construction and copying of machines creates parts that are not the product of human intention towards an end, yet which make a significant contribution to the machine. Granting that there are cases where intention cannot be used as

[12] For a more timely example, consider that bees do not distinguish between flowers coated in insecticides and those without. This has apparently resulted in a decline in the population of bees, yet hairiness serves the same purpose for flowers regardless.

the explanation of extrinsic teleology, we must look elsewhere for the source of purpose. In organisms, we can ascribe teleological characteristics on the basis of the relationship between the provider and recipient of some effect—a purpose is intrinsic when bearer and benefactor are the same, extrinsic when they are different. With organisms, this benefit is best thought of in terms of increases in reproductive success or persistence. Likewise, I argue, we can analyse benefit to machines in terms of increases in production or persistence (Ch.6). Thus we can establish that some machines have intrinsic teleology by showing that they can be the beneficiaries of their purposes; that they can have their own ends or, at least, ends separate from those of our intentions.

2.5.1 What is the function of fortuitous accidents?

Fortuitous accidents in the design of machines can have unintended effects. Consider the following hypothetical origin story: a designer is creating a machine M for the purpose of doing X. For instance, a chip for a radio. In the process of designing the chip, the designer reasons about what the chip must do overall, X, and how specific parts should contribute. The designer has some understanding of the interaction between parts A-Z, so assembles them into M according to a design that is intended to bring about X. As it turns out, M does X, and so the designer rests content and begins mass-manufacturing copies of M.

Nonetheless—despite a designer's understanding—some machines might include parts or effects that were not included in the design nor intended to contribute to the operation of the machine. As Lewens (2013) notes, design methodologies can result in machines that are not easily comprehended by functional decomposition: "innovators may have unwittingly failed to reduce [interactions] that, unknown to them, make an important contribution to the overall operation of the artifact" [ibid]. Unintended effects can make a positive contribution to the capacity of M to X. Suppose further that within M there are two parts A and B. These parts were intended to have effects F_A and F_B that were intended to contribute to the capacity of M to X[13]. Now, suppose that the interaction A:B between A and B has some other effect $F_{A:B}$ that is unknown to the designer. Moreover, equally unknown is that without $F_{A:B}$ M would not X, i.e. A:B is a cause of M's capacity to X.

Such cases are not negligible; examples are growing in number and significance. On the extreme end, we can imagine a serendipitous machine for which the designer knows nothing about how individual parts contribute to the overall capacity X. Artificial intelligence

[13]The example is intentionally simple, though I am told by electrical engineers that they believe the case described is a common one. We might equally supposed that A and B are part-types, instead of part-tokens; that X is an effect that may obtain to varying degrees; that F_A and F_B are a collection of related effects, each contributing partially to X. None of these change the upshot of the example.

machines now loom large as examples of machines that become more mysterious as they grow more powerful, as their intelligence grows in proportion to our inability to understand their finer workings. In an article for the MIT technology review[14] entitled "The Dark Secret at the Heart of AI", Will Knight points out that many AI are "so complicated that even the engineers who designed it may struggle to isolate the reason for any single action." Once an initially general purpose AI has learned or been trained, the purpose for most internal effects are not only unknown and unintended, but not understandable. Though it is, for now, still clear that the AIs as a whole have a teleology that is imposed by the designer—or trainer—this is not so for the particular effects of their parts.

What should we say about the purpose of parts like A:B? Intuitions about whether fortuitous accidents have 'functions' are split. Wright (1973) argues that they are accidents, Kitcher (1993) thinks of them as 'functions' without intention. I will take no stand on the correct intuition behind the word, but will show that indeed the etiological and teleological accounts presented here do apply to such parts.

To begin with, one sort of functional analysis involves merely identifying the effect that each part has on a containing system and ascribing each a function / purpose if it contributes causally to the overall effect of a system. These are so called Causal Role (CR) or Cummins functions (Cummins 1975). Here, we can say that the CR-function of A:B in M is to $F_{A:B}$ in virtue of the fact that $F_{A:B}$ contributes to the capacity of M to X. Minimally, unintended effects do have CR-functions, yet these alone will not establish a difference with organisms—both sorts of entities can be subjected to a CR analysis. Something more is required to establish whether fortuitous accidents should have purposes in a more richly intrinsically teleological sense.

Consider the etiological case first. Unintended effects of fortuitous accidents are like dairy cows: they have both a rational and natural etiology. Though A:B exists in part by virtue of the rational intervention on nature by the designer, the unintended effect $F_{A:B}$ does not. $F_{A:B}$ instead owes its existence to the naturally arising effects of the parts A and B. Though the occurrence of A:B is in part due to the rational and intentional manufacture of machines M to do X, it is nonetheless also due in part to the natural contribution of $F_{A:B}$ to X. Here, intended and actual effects come apart. While the rational etiology of M involved only the intended effects F_A and F_B, the natural etiology also involves the actual effect $F_{A:B}$. In Wright's (1973) terms, were it not for $F_{A:B}$, A:B and the containing M would not have been consciously-selected by the designer, even though it was not a reasoned consequence of the designer's intentions. If we ask "Why is A:B in M?" while concentrating only on intentions and design, the answer will be "A:B is there by accident." But if we look at all of

[14] https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (August 3, 2018)

the causes for A:B's presence in M, we can answer that "A:B is in M because it does $F_{A:B}$." An etiological analysis of some machines should describe them as both artifact and natural entity.

To provide a teleological account we require an analysis of the relationship between purpose and benefit. When we say that the purpose (or function) of X is to F, there is an implicit third relatum: the beneficiary[15]. Really, we should say the function of X is to F for Y. Notice how the identity of beneficiaries determines the variety of teleology. As Nicholson (2013) says, rightly, whether something's *telos* is intrinsic or extrinsic depends on which 'agent' it is of use or value to and on whether it 'acts' on its own behalf. When all the parts of some entity Y are functional for Y then Y has a strictly intrinsic teleology. Above I attempted to show that many organisms contain parts that are functional for an external entity and thus that organisms are not strictly intrinsically teleological. Likewise when all the parts of an entity Y are functional for some other entity Z, then Y is strictly extrinsically teleological. So to argue that some machines contain parts that are functional for those machines further requires an analysis of what it means for a machine to be a beneficiary of a function of one of its parts.

One way to define benefits to entities is in terms of fitness, interpreted as (at least, actual or potential) increases to survival and/or reproduction. Considering fitness as a matter of survival, we can analyze "The function of P is to F for Y" as "The effect F of P contributes to the persistence of Y" (Doolittle 2014; Bouchard 2008). Indeed, if P is a part of Y then F would, in keeping with the organizational account that is supposed to underpin Nicholson's (2013) analysis, be a capacity of self-maintenance. On the side of reproduction, we can analyse "the function of P is to F for Y" as "the effect F of P contributes to an increase in the number of Y". How F affects an increase in Y will surely differ with machines, but the same teleological analysis can be provided. The effect F of P is its purpose, and it is directed towards Y, it is the telos of Y, provided it affects an increase in persistence or increases in number of Y. We can consider machines as beneficiaries in both of these senses.[16]

What it means for a machine M to be the beneficiary of a function F of one of its parts P, is that the effect F of P in M contributes to an (actual or potential) increase in the number or

[15]There are many more relata of course. On CR functional analysis we really ought to say something of the form "The function of X is to Y in Z on analysis A for O at time T", but these further relativizations can be set aside here (but see Cummins 1975).

[16]Recall that at the beginning of this chapter I endorsed the SE account of function—I still do here. This section is an engagement with the consequences of the organizational or Kantian view of functions, as advocated by Nicholson (2013), qua MOD, not an endorsement of this approach to teleology in biology.

persistence of M.[17] Beginning with increases in number, since $F_{A:B}$ contributes to the capacity of M to X, and X is (by hypothesis) part of the reason that M's are manufactured, M's *benefit* from the effect $F_{A:B}$ of their constituent A:B's. Thus the copying of machines on the basis of their overall capacities for us ensures that parts that contribute to their capacities also have functions for machines—machines benefit from the purposes of their parts. Likewise for persistence. There are indeed machines that have effects on their own persistence or durability, so we need only imagine that $F_{A:B}$ contributes to some such effect X—such as self-cleaning, or self-defragmentation (Ch.3). In that case M would be, in part, the beneficiary of $F_{A:B}$. This is all that is required to say that M has some part with intrinsic teleology.

What makes cases like the bee-flower symbiosis and unintended machine parts so perplexing is the assumption that intrinsic and extrinsic teleology are mutually exclusive. They are not: one and the same entity can have purpose arising from within and imposed from the outside. Perhaps ideal organisms would do nothing that they themselves were not the beneficiaries of, but real organisms are often co-opted by others, enslaved, or cheat in former mutualisms. Perhaps ideal machines serve no other purpose than what they were intended to perform for our benefit, but real design processes can create machines with purposes that are not imposed by designers but arose naturally. Just as with natural and rational etiology, intrinsic and extrinsic teleology fail as an MOD.

2.6 Conclusion: Spectra of Purposes

Adopting an etiological view of the machine organism distinction does not provide a mutually exclusive MOD since one and the same entity can possess both a rational and natural etiology. Coupled with the possibility that synthetic biology might generate intuitively organismal entities with a totally artificial history, this suggests the plausibility of an alternative, more Kantian and teleological, view of the MOD. Nonetheless, cases abound where one and the same entity or part can have both an intrinsic and extrinsic teleology. Symbiotic interactions between (rather than within) individuals, non-rational or unintentional designs, and the blind copying of fortuitous machines all speak against the view that organisms operate towards their own ends and that machines operate towards ours. The distinction between an artificial and natural history or between intrinsic and extrinsic teleology may sometimes be good

[17] I am here glossing over a type-token distinction that is surely applicable: the difference between tokens and types of machine M. So long as some type-token distinction can be made, on the basis of contributions to copying similar machines, then we can safely exclude this detail from the present argument.

heuristics for talking separately about organisms and machines as classes, but these are not their characteristic properties.

This suggests another view of the teleology and etiology of organisms and machines. Entities sit on a spectrum of purposes; they can be classified according to their coordinates on a teleological-etiological plane. Familiar organisms and machines might sit at opposing corners (Fig.2.1, top-right and bottom-left). Although, on either of these scales, other machines and organisms will sit at positions between the extremes. The following figure sketches how this might look.

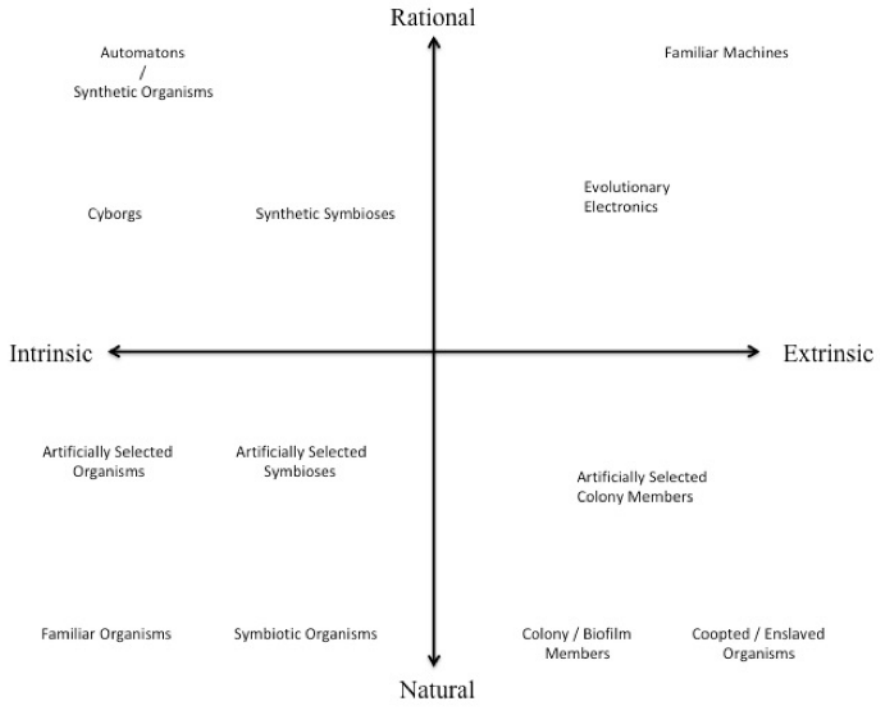


Fig. 2.1 Spectrums of rational-natural etiology opposed to intrinsic-extrinsic teleology. Positioning of examples within broad categories is illustrative.

Plainly, the historical explanation for the traits of an entity is not restricted to the history of entities of its type, but may involve reference to the functions or intentions of other entities. Sometimes the traits of an entity will be explained by the functions of other entities; sometimes the functions that are *for* one entity will be *of* another entity. If we wanted to provide an historical explanation of the traits of any entity that lives symbiotically, has been artificially selected, enslaved, bred, cultivated, otherwise functionally-coopted (by human or non-human entities), genetically, surgically, or technologically modified, then this explanation will involve a spectrum of purposes. Organisms may have purposes that

are directed outside and operating—deleteriously or not—towards the ends of what might otherwise be considered the (technological) environment. In a phrase with post-modern appeal, for these stranger sorts of living and technological things we need a cyborg teleology (§ 7.2.2).

Now that we have considered the content of teleological judgements about machines (their telos or direction), we should consider their source (autonomy or self-organization). The following chapter addresses views of organization and autonomy of organisms, their apparent lack in machines, and how these bear on the MOD.

Chapter 3

Autonomy and the Machine Organism Distinction

Abstract

In this chapter I argue against the views of Maturana and Varela (1980) and Moreno and Mossio (2015) that autopoiesis or autonomy is the MOD. I argue against this by showing that both machines and organisms demonstrate forms of marginal or incomplete autonomy. To show this, I analyse autonomy into a number of interrelated "self-directed capacities" (e.g., self-organization, self-synthesis-of-adenine, etc.). There are organisms that are dependent on others to survive, organisms that lack certain self-directed capacities; there are machines that are capable of persistence and some reproductive operations without human intervention, machines that possess certain self-directed capacities. Together this shows that autonomy is one of the characteristics on which organisms and machines vary, not their characteristic difference.

3.1 Introduction

The previous chapter addressed the teleological content of a Kantian view of organisms, as defended by Nicholson (2013). I argued that the direction of ends, whether their *telos* is self- or other-directed, does not fundamentally distinguish machines from organisms. To do that it was necessary to background the sources of teleological properties somewhat, in an attempt to isolate any special role that is played by the content of teleological claims. I argued that if the source of teleological properties is etiological, a relationship to rational or natural history, then this will not serve as an MOD. This chapter considers the source and backgrounds the

content of teleological judgements. Specifically, it addressed the idea, hinted at by Nicholson (2013), though addressed more directly in Maturana and Varela (1980) and Moreno and Mossio (2010; 2015), that the source of teleology in organisms is the distinctive form of Kantian reciprocal causation between their parts.

Maturana and Varela (1980) address this in their theory of autopoiesis; Moreno and Mossio (2010; 2015) as the theory of biological autonomy, or organization. What these perspectives share is an emphasis on the autonomy of organisms, their self-organization and self-maintenance, and the conclusion that machines and organisms can be distinguished on the basis of these properties. In this chapter I argue that an appreciation of microbiology and contemporary robotics shows both that autonomy is not the characteristic feature of organisms and that it does not provide an MOD.

This chapter consists of two main parts: this first (§ 3.2-3) is a conceptual analysis of biological autonomy, the second (§ 3.4-5) is a naturalistic catalogue or survey of marginal or borderline autonomous systems. In § 3.2-3 I will explicate the views of autopoiesis offered by Maturana and Varela (3.2.1) and how these developed into the view of biological autonomy offered by Mossio and Moreno (§ 3.2.2). Both of these views are often obscured by their highly abstract and variable terminology, but a plausible reading sees them as offering up an account of organisms on the basis of some intuitive criteria for autonomy. I then argue that autonomy can be analyzed further into a large number of self-directed capacities (§ 3.3, e.g. that self-organization includes self-synthesis-of-required-chemicals).

Once we see autonomy as consisting of a large number of self-directed capacities, it is clear that many organisms lack forms of autonomy, and many machines possess them (§ 3.4). Autonomy obtains in a balance with its antithesis, here called 'allonomy'. There are marginal or borderline autonomous systems in both biology (§ 3.4.1) and technology (§ 3.4.2). In § 3.4.1 I examine cases where organisms must acquire energy, synthesize their parts, or reproduce with the aid of others. § 3.4.2 examines cases where robotics have some self-organizing, self-maintaining and self-reproducing capacities. These cases amass evidence that autonomy does not establish a fundamental difference between machines and organisms, but another spectrum along which entities of both types *predominantly* occupy intermediate positions (§ 3.5).

3.2 Autonomy and the Machine–Organism Distinction

3.2.1 Autopoiesis and Allopoiesis: Maturana and Varela

> Autonomy appears so obviously an essential feature of living systems that whenever something is observed that seems to have it, the naïve approach is to deem it alive. Yet, autonomy...seems so far the most elusive of their properties. —Maturana and Varela (1972)

In their book *Autopoiesis* (1972), Maturana and Varela present an account of organisms based on autonomous organization. We will soon see how this view establishes an MOD. However, we should first note what they claim is *not* involved. They distinguish their view from those involving teleology or purpose (Ch.2), which they claim are "obviously naïve" ([ibid] p.77), and from accounts that stress the importance of a particular composition. Varela et al. (1974) explicitly state that "reproduction and evolution are not constitutive features of the living organization and the properties of a unity cannot be accounted for only through accounting for the properties of its components."

To understand the MOD provided by autopoiesis, we should begin with the account of "machines" on offer.

> [M]achines are unities...they are made of components that are characterized by certain properties capable of satisfying certain relations that determine in the unity the interactions and transformations of these same components... — Maturana and Varela (1972)

The foregrounding of 'unity' is an endorsement of holism. Machines are components that work together and should be understood together, unlike some other transforming collections, like junk-piles. This very general account of machines is given in order to make a further distinction between two types of machine: allopoietic and autopoietic. These are derived from *allos* meaning "different" or "other", *auto* "self", and *poietic* "production". Autopoietic machines produce themselves; allopoietic machines produce (and are produced by) something other. It is through autopoiesis that they provide an account of organisms. According to Maturana and Varela (1972), 'autopoiesis' is "the organization which makes a living system a whole, autonomous, unity that is alive".

They do provide a definition of 'autopoietic machine' ([ibid] p.79), although their definition is somewhat obscure[1]. Much of *Autopoiesis* can be seen as repeated attempts to give similar definitions. What is essential in their definitions, and has stood the test of time, are

[1] One such definition appears in full in the following section, for later comparison.

the following features. Autopoietic machines are organized such that (a) their components interact to regenerate the same sorts of components, and (b) they maintain themselves, as a homeostatic whole, in the face of external perturbations. I will call (a) 'self-organization' and (b) 'self-maintenance'.

The concept of allopoiesis is intended to align with that of more familiar machines. They provide the following definition for the allopoietic: "machines that have as a product of their function something different from themselves, as in a car" ([ibid], also see Varela et al. 1974, p.8). Here it is some functional effect like locomotion, and not fumes, that is the intended "product". More clearly, the products of machines on a factory assembly line are typically not the assemblers themselves. Maturana and Varela (1972) use 'allopoietic' as a catch-all: including most familiar machines and designating anything not-autopoietic—a machine is allopoietic if it is not self-organizing and not self-maintaining.

Here is the MOD: Organisms are a type of machine, the autopoietic type, whereas familiar machines are allopoietic. Organisms are autonomously maintaining their organization; familiar machines are non-autonomously producing something else. Varela et al. (1972) are direct,

> Allopoietic systems are by constitution non-autonomous insofar as their realization and permanence as unities is not related to their operation.—[ibid] p.8

The view offered by Maturana and Varela is the skeleton of a complete view. They only claim that unity is *somehow* maintained and organized, without explaining "unity" or how it is achieved. Mossio and Moreno (2010), described in the following section, go further towards filling these lacunae.

3.2.2 Self-Organization: Mossio and Moreno

Mossio and Moreno (2010; 2015) offer an account of biological autonomy that stems from the work of Maturana and Varela (1972).[2] Their view drops the use of 'autopoiesis', but retains emphasis on organization, maintenance, and regeneration. It also adds explicit reference to additional autonomous capacities and requirements for energy. As well as relationships of production, this newer organizational account emphasizes thermodynamic relationships. Beginning with the account of 'machines',

[2]Unfortunately, as Garson (2017) points out about the closely related organizational theory of functions, "the details and terminology vary somewhat in different publications" (c.f. Mossio et al. 2009; Mossio and Moreno 2015). I will follow the details and terminology offered in Moreno and Mossio (2010) and Militello and Moreno (2018).

> [W]e define a machine as a meta-stable structure, which can persist in thermodynamic equilibrium, consisting of a number of functionally interdependent parts that constrain an energy flow to do work and perform a systematic function.—Militello and Moreno (2018) p.35

As with autopoiesis, we can see the account of organisms as carving out a special class of machines. This time, depending on which work is done and how this affects existing constraints within the machine. The essential ingredients of the account of organismal autonomy are tripartite, including (1) self-maintenance, (2) organizational closure, and (3) reproduction and regulation. These will be described in turn.

(1) Self-Maintenance—To understand what Moreno and Mossio add to Maturana and Varela, it is now worth considering the latter's definition of autopoietic machines.

> An autopoietic machine is organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network.—Maturana and Varela (1972) p.79

Moreno and Mossio (2010) take (i) in the above definition and replace the emphasis on components with that of constraints and add explicit mention of the role of energetic transformations that was absent in Maturana and Varela. Self-maintenance "results from the action of local constraints", and a system is self-maintaining when "it is able to use its work to re-generate at least some of the constraints that make work possible". Put another way, self-maintenance consists of a relatively simple feedback loop where work generates a constraint that positively effects the ability to do precisely that kind of work. Self-maintenance is essentially maintenance via partially reciprocal causation. The authors acknowledge this quality in many non-living systems, e.g. candle-flames.

(2) Organizational Closure—The move to organizational closure involves more complex, holistic forms of self-maintenance. Here is the connection to Kantian teleology. Mossio and Moreno (2010) trace this idea back to the reciprocal causation advocated in Kant's *Critique of the Power of Judgement*. Each organ is producing and being produced by all others. They refer to this as an organism's ability to "self-organize", where the limit of self-organization is, at the high end, organizational *closure*. The term 'closure' itself comes from Varela (1974), but I think it easiest to trace their formulation back to part (ii) of the definition of autopoietic machines, given above. Mossio and Moreno (2010) write,

> [B]iological systems are able to maintain themselves by constituting a web of structures exerting mutual constraining actions on their boundary conditions, such that the whole web is collectively self-maintaining. The mutual dependence between a set of constraints is what we call an organizational closure. –[ibid] p.276

Besides the diverging terminology between Maturana and Varela (1972) and Mossio and Moreno (2011) for apparently similar notions—between 'web' and 'network', or 'boundary conditions' and 'topological domain'—we are dealing in approximately the same idea: a more complete form of biological autonomy requires a collection of interdependent self-maintaining systems. Both are elaborations on Kant's view of reciprocal causation. Reciprocal causation may obtain between pairs of parts in achievement of self-maintenance, but organization closure requires "mutual dependence" [ibid],[3] where reciprocal causation obtains between a "web of elements" [ibid].[4] Moreover, this is supposed to be a "fundamental property" or "essential mark" of the biological [ibid]. While candles may be self-maintaining in a simple sense, their organization is "open" in that it requires the constant input of wax that is not affected by any constraint of the candle itself.

(3) Control and Reproductive Mechanisms—Moreno and Mossio (2010) say that (1) and (2) are necessary but not sufficient for being an organism: "all organisms are organizationally closed systems, [but] we do not conclude that any organizationally closed system is a (fully-fledged) organism" [ibid]. Although organizational closure is apparently not a sufficient condition for being an organism, their association between organizational closure and organisms is nonetheless quite strong. They say that (1) and (2) "[make] the crucial transition between the physiochemical and the biological domain", that they "constitute a fundamental property of the distinctive causal regime at work in biological organisms" and that (2) can be taken as an "essential mark of living organisms". Evidently the transition to fully-fledged organisms is slight. To that effect they offer a final ingredient to their view of organismality: "reliable reproduction and control mechanisms". The addition of control and reproductive mechanisms is to "enable the organism to respond adequately and adaptively to external perturbations". While the inclusion of reproduction is a break with Maturana and Varela (1972), the requirement for control and adequate response to "external perturbations" is clearly in line with (b), (ii) and with self-maintenance generally.

Mossio and Moreno (2010) make two further restrictions on their view. They point out that organizational closure is partial and context dependent. In other words, (1) holds "only if

[3]The approximate equivalent in Kant (2000) is his technical use of 'perfection'.
[4]Kant (2000) says "manifold elements".

a given set of independent boundary conditions is also fulfilled" and (2) requires that "at least some, but not all, the constraints under which a system can exist have a relation of mutual causal dependence". These are biologically realistic constraints on a theory of organisms, but they each leave open the possibility that the theory of autonomy on offer might be unable to provide an MOD. Partiality and context dependence, as well as being relativistic, also admit difference in degree. The ability for autonomy so construed to provide an MOD would then rest on whether there are contexts in which some machine might demonstrate self-maintenance, closure, etc., to a degree comparable to the partiality of some organism. This is the topic of § 3.4.

3.3 Analyzing Autonomy

3.3.1 Self-Directed Capacities

Discussions of autonomy involve its analysis into a number of capacities. These include self-organization and self-maintenance (self-persistence), self-generation, self-regeneration, self-constraint, self-repair, self-production, and self-reproduction. Preference for specific synonyms differ between authors and sometimes publications. To this list we should add self-regulation (Walsh 2015) self-development (when not covered by others), self-reorganization (for metamorphic organisms), self-destruction (for apoptosis or, potentially, forms of "programmed aging"), and perhaps even self-evolution (to account for recently discovered "evolvability" mechanisms (Janković and Ćirković 2016; Koonin and Wolf 2012)). Each of these involve some capacity that is possessed by an organism (a particular "self") and directed towards it. I will refer to all of these as *broad self-directed capacities*.

The working assumption in theories of autonomy is that it requires joint possession of some number of self-directed capacities, and that organisms are paradigmatic examples of such autonomous systems. The distinction with machines is then provided negatively: machines, lacking self-directed capacities, are *non-autonomous*. Machines may be organized or do some organizational work, but they are *not self-organizing*; they may be maintained or maintain something, but they are not self-maintaining; and they may be produced (and reproduced) or produce something, but they are not self-reproducing. We, not machines, are the reason that machines are produced, maintained and organized.[5] When machines do seem to possess a modicum of these capacities, they are considered marginal examples at best.

[5]Of course, it may be that some machine that repairs another one, but that chain will (so the working assumption goes) quickly lead back to us.

Clearly not all paradigmatic organisms are capable of self-destruction, this being a uniquely multicellular invention; some organisms are too ephemeral to heal large non-lethal wounds; most organisms do not regrow severed limbs. So which particular set of self-directed capacities make a system autonomous? Answers differ. Nicholson (2013) lists self-organization, self-reproduction, self-preservation, and self-regeneration, as well as (more difficultly aligned with the above examples) self-formation, and self-restitution. Mossio and Moreno (2010) instead focus on "organizational self-maintenance", but cite the need for forms of self-regulation and self-reproduction. Maturana and Varela (1980) explicitly deny that self-reproduction is essential, focusing instead on self-regeneration, self-maintenance and self-organization.

In the absence of consensus I will focus on self-organization, self-maintenance and self-reproduction. These broad self-directed capacities are not wholly distinct, but are nearly universally shared in analyses of autonomy.[6] They also most closely match the biologically important notions of metabolism, homeostasis and reproduction. In the following section I argue that these broad capacities should be analyzed into a much larger list of refined capacities.

3.3.2 The Multiplicity of Self-Directed Capacities

Autonomy can be analysed first at a coarse grain into what I have called the broad self-directed capacities. But that grain of analysis is not sufficient, I argue, to appreciate the diversity in autonomy. For that, we need a more fine-grained analysis of autonomy, into a collection of refined self-directed capacities. Mending a broken heart is different from initiating cardiac tissue repair mechanisms. Both are forms of self-maintenance, but the latter is self-maintenance considered at a more fine physiological grain of analysis. This section argues that we must spell out a much larger multiplicity of such capacities, for each of the broad self-directed capacities, if we are to properly understand diversity in autonomy.

Consider self-organization. Self-organization is, in part, about the capacity that an entity has to recreate its components (or constraints). One of the primary endeavours of biochemistry and physiology is the investigation of precisely how certain components interact to accomplish this. These investigations usually aim to identify a particular metabolic pathway or key reaction that has a given component (constraint) as one of its products. We can look at these empirical investigations as providing us with a basis for an analysis of

[6]For an example of how they are not distinct: the synthesis of "heat-shock proteins" during adjustment to increasing environmental temperature might be considered an example of self-organization, because they assist in folding of other proteins, or self-maintenance, because they compensate for perturbations. Likewise, synthesis of platelets is a form of self-organization, but platelets are required for repair of blood vessels.

self-organization into a number of more refined self-directed capacities, interrelated self-directed capacities of self-synthesis and self-degradation. Each metabolic pathway or reaction provides a capacity (to synthesize or degrade something relevant to organization), possessed by a self, whose product may be a component (or constraint) in other pathways. Analysing the broad self-directed capacities allows us to do two things that were previously unavailable to us.

Firstly, this analysis of self-directed capacities allows us to refine the question about which capacities are required for autonomy. Since autonomy is defined in terms of self-directed capacities, we see that autonomy is likewise analysable into a multiplicity of more refined forms of autonomy. In the following sections this will be important, since it allows us to talk about borderline, partial, or marginal cases of autonomy at a finer grain than possible with the broad capacities alone. We can then supply more nuanced judgements than are permitted with only a distinction between "fully-fledged" and "not fully-fledged" autonomy.

Secondly, analysis of the broad self-directed capacities allows us to consider their interdependencies on a more individual basis. Mossio and Moreno (2010) are inclined to treat autonomous capacities as "a distinct level of causation, operating in addition to the physiochemical laws" [ibid, p.270], while Maturana and Varela (1972) say that autopoietic machines are "defined as a unity". I take these to be claims about the interdependence of self-directed capacities, and this reading is upheld by the notion that more fully-fledged autonomy requires organization-closure. Moreover, if these *prima facie* metaphysical claims are best interpreted as about causal interdependence among self-directed capacities, then this should not be established *a priori*. Considering a more refined list of self-directed capacities allows us to substitute some metaphysical problems of biological causation with empirical ones. What matters is that we can consider a given self-directed capacity, e.g. 'an entity's capacity to synthesize its own alpha-tubulin', and see if a given organism is capable of doing this independently from another capacity or another organism. It is an empirical question whether refined forms of self-maintenance are bundled together into a distinct-level, unity, or whether they rather enjoy some forms of physiological independence.

Even certain broad self-directed capacities enjoy forms of physiochemical independence at that coarser grain of analysis. That self-maintenance does not depend on self-reproduction was a key motivation for Maturana and Varela (1972) for excluding the latter from their account of autopoiesis. Neuter animals, eunuchs, post-reproductive females, and the last surviving members of sexual species are autonomously self-organizing and self-maintaining, but lack the possibility of self-reproduction. The same kinds of causal independence likewise obtains for some of the refined self-directed capacities. Patients suffering from phenylketonuria (PKU) have a lost or reduced capacity for the self-degradation-of-phenylalanine, due to a

mutated version of the enzyme phenylalanine hydroxylase (Penrose 1951; Scriver and Waters 1999). Patients suffering from PKU have a buildup of phenylalanine, but not of other amino acids, indicating the self-degradation-of-phenylalanine is independent of self-degradation-of-other-amino-acids. On the other hand, phenylalanine hydroxylase is involved in the synthesis of tyrosine, so self-synthesis-of-tyrosine is dependent on self-degradation-of-phenylalanine (Payne and Loomis 2006). Tyrosine can only be self-synthesized when phenylalanine can be self-degraded, in humans.

Evidently these analyses of autonomy have overstated some of the holism, collective self-maintenance, closure, or unity of self-directed capacities, while their multiplicity has been understated. The following section argues that an appreciation of the diversity in self-directed capacities among organisms works against treating autonomy as their characteristic feature, and diversity in the self-directed capacities of machines works against the MOD.

3.4 Diversity in Self-Directed Capacities

Before turning to counterexamples to MODs based on autonomy (4.1-2), we should be clear about what autonomy contrasts with. I have been discussing autonomy one-sidedly, as if entities were either autonomous, less autonomous, or non-autonomous. Like the autopoiesis-allopoiesis distinction, autonomy has an opposite. Non-autonomy is only its negation; anti-autonomy should involve reference to *other selves*. A rock is non-autonomous, but need not have any particular relationship to selves.

What the opposite of autonomy is named seems to differ depending on what particular broad or refined self-directed capacity is at issue. For instance, Wernegreen (2012) discusses interactions between symbionts and their hosts as sorts of *integration*; Godfrey-Smith (2013, p.70) discusses things that "reproduce with the aid of much external machinery" as *scaffolded*; and Queller and Strassmann (2009) discuss social organisms in terms of *cooperation*. We might also think of anti-autonomous relationships cybernetically, as involving dependence or forms of external *control* (R. Kline 2009). So for instance, in Godfrey-Smith's vernacular, we should say that 'autonomous reproduction' is to be contrasted with 'scaffolded reproduction', while Wernegreen would discuss this in terms of 'reproductive integration'. This is a confusing way to discuss contrasts with autonomy. I will be specific about precisely which capacities are at issue, so will use the more theoretically neutral and directly opposite '*allonomy*'[7].

[7]This term appears in Hirst (2008) in direct connection with the issue of autopoiesis and autonomy, though therein only designates a rigid non-autonomy supposedly possessed by machines.

We should not confuse allonomy with dependence on the abiotic environment. Of course, every organism depends on its environment, but allonomy should be a more richly biotic notion. Mossio et al. (2009) are correct to point out that "self-maintenance has nothing to do with 'independence' from the environment or external conditions." Mossio and Moreno (2015) hold a similar view about self-organization. They allow for dependence on the environment because of the need for influx of energy and material: "Autonomous systems are...organizationally closed and thermodynamically open" [ibid, p.6]. But not all dependence on the environment is abiotic and thermodynamic, some involves being *organizationally open*. If *self*-maintenance has to do with independence from anything, it should be independence from *other*-maintenance, and likewise for self-organization and other-organization. Designations of allonomy should be made with respect to dependence on the *biotic* environment.

I will also distinguish between what might be called material and mechanistic forms of allonomy[8]. The distinction is helpful because some forms of allonomy involve a reliance on materials from others, while some involve an outsourcing of mechanisms to others. These will often be intertwined in real cases but can mark out interesting differences, discussed below. While some reliance on others for material conditions is perhaps the norm in biology, there are more complicated mechanistic forms of allonomy in organisms that seem to be neglected in discussions of their autonomy.

The natural world, both biological and technological, is diverse, and part of this diversity is manifest in differences on a spectrum from autonomy to allonomy. The following sections are largely an empirical catalogue of allonomy in organisms and autonomy in machines. I argue that the majority of organisms lack some of the central features of autonomy, or possess them only marginally (4.1). Likewise, many machines seem to have autonomous features, some spanning each of the broad self-directed capacities (4.2). Of course, there are different ways to interpret this diversity. It might be taken to show that "fully fledged" organisms are quite rare. I think it is best interpreted as showing that autonomy fails as an analysis of organisms, since many are not fully autonomous, and as an MOD, since the difference with machines then amounts to a contingent difference in degree on the autonomy–allonomy spectrum.

[8]Here these terms are meant in their most metaphysically neutral senses and should not be taken to support a neo-mechanistic worldview. If it turns out that a process ontology is the most appropriate to biological phenomena, then this is a regrettable terminological choice, perhaps better replaced with a distinction between 'substantial' and 'processual' allonomy (see Ch.4).

3.4.1 Organisms that Lack Specific Self-Directed Capacities

For a self-directed capacity lacking in an organism, lets first examine chemosynthesis and the self-directed capacities of energy acquisition. At the extreme end of parameters for life there are organisms capable of complete energetic autonomy. This title goes to some of the chemolithoautotrophs and the photolithoautotrophs, organisms that use inorganic chemicals (chemo) or light energy (photo), inorganic electron donors (litho) and inorganic carbon (auto) as resources (troph). Ehrich et al. (1995) report a particularly interesting case of an obligate chemolithoautotroph, an organism that cannot grow in the presence of other organic materials, found growing on a rusty pipe in Russia. This is perhaps the pinnacle of autonomy—to subsist on rust, light, and air—but it is not the norm. Autopoiesis and organizational closure might work as theories of the autonomy of lithoautotrophs (see Jones et al. 2011, p.179). But even the lithoautotrophs are not entirely autonomous, some such microorganisms depend on others to acquire energy sources.

In the metabolic classification used above, each part of the name specifies a requirement for metabolism. The order is – energy source – electron source – carbon source, and the possible combinations are chemo/photo – litho/organo – auto/hetero/mixo. The chemo/photo distinction separates organisms that gain energy from chemicals or light, yet the discovery of fungi around the reactor at Chernobyl leads some to also include radiotrophy (Karpenko et al. 2006; Dadachova and Casadevall 2008). The prefix 'litho' means that an organism obtains electrons for metabolism from "inorganic" sources, here having the precise meaning of "not containing carbon", while 'organo' means that electrons are obtained from organic sources. The auto/hetero distinction refers to whether the organism obtains carbon from atmospheric sources or from consuming parts of other organisms. Given this classification, it is apparent that some combinations of metabolic sources would allow an organism to survive and reproduce without the presence of others (or their parts) provided a sufficient abiotic environment. Lithoautotrophs can be "at the base" of an ecosystem or food chain, acting as "primary" producers of organic compounds for others (Cavanaugh 1994). But note that any of these metabolic sources can and sometimes are obtained from other organisms. Both carbon-containing ("organic") and carbon-non-containing ("inorganic") energy sources and electron sources are often obtained from other living organisms (by heterotrophy).

Cavanaugh (1994) describes the case of symbiotic interaction between sulfur-oxidizing chemoautotrophic bacteria and marine invertebrates found both in deep-sea hydrothermal vents and coastal sediments. The giant tube worm *Riftia pachyptila*, for instance, completely lacks a mouth and digestive tract. Instead of digesting organic matter from its environment, it relies entirely on the biomass generated by endosymbiotic sulfur-oxidizing chemoautotrophs. This is quite a significant case of energy and carbon harvesting from chemoautotrophes;

giant tube worms can grow up to 2m in length. Likely speaking of coral-algae symbioses and drawing inspiration from Margulis and Fisher (1991), Cavanaugh (1994) proposes that "invertebrates "become" photosynthetic via symbiosis with algae" [ibid p.79]. We might likewise say that invertebrates become chemosynthetic via symbioses with sulphur-oxidizing bacteria.

What the bacteria gain in return from their host is access to both an energy and electron source. The blood of *R. pachyptila* contains a specialized form of hemoglobin capable of transporting both sulphide and oxygen. Combined with carbon dioxide from the ambient environment, this is sufficient for autotrophic growth. Interestingly, Cavanaugh (1994) speculates that this relationship evolved because the concentrations of sulphide and oxygen in seawater are often inversely proportional, requiring most sulphur-oxidizers to attach to a mobile surface that passes between high-sulphide and high-oxygen regions—a form of mechanistic allonomy for an organism that lacks a sufficient capacity to self-transport and self-feed. In order to survive this bacterial endosymbiont depends on a giant tube worm to provide it with sufficient access to energy. Even in "primary producers" we find symbiotic interactions that produce forms of allonomy.

This is a clear case of other-maintenance of an organism's energetic capacity: a form of mechanistic and material allonomy that is ubiquitous and sitting at the base of many ecosystems. Chemosynthetic symbioses provide diverse cases of marginally autonomous organisms, cases where some of the refined capacities of organization are not entirely self-directed. These cases satisfy part, but not all, of Mossio and Moreno's (2010) account of self-maintenance. By providing organic materials to the worm a sulfur-oxidizer "is able to use its work to re-generate at least some of the constraints that make work possible". But these constraints are not "local", some residing in the worm and others in bacteria. Moreover, self-reproduction and self-synthesis-of-organic-material also come apart. The bacterial symbionts of *R. pachyptila* are acquired 'horizontally' (from the environment) and largely retain their free-living self-reproductive capacity. The worms themselves are dependent on the energy and material provided by their temporary symbionts to reproduce at all—giant tube worms have material reproductive allonomy. On the other hand, the chemosynthetic endosymbionts of clams are transmitted 'vertically' (from parent generations), are not free-living and have lost the capacity to self-repair and self-recombine-DNA (Dubilier et al. 2008).

For another example, consider the self-directed capacities of biosynthesis and biodegradation. Most organisms cannot synthesize everything they need to survive and they cannot degrade everything that hinders their survival. Up to 99% of microbes cannot be cultured in isolation (Pham and Kim 2012). This "unculturability" of microbial life has received growing attention from scientists and philosophers, interested both in devising new culture methods

and explaining how this situation evolved (Dupré and O'Malley 2009; O'Malley 2014; Wolf and Koonin 2013; Brunet and Doolittle 2018; Pande and Kost 2017). Importantly for us, this phenomena shows just how common it is that prima facie organisms—microbial cells with clear "boundary conditions" and a "topology"—are unable to survive by themselves.

One explanation for unculturability is syntrophy or "cross-feeding": a relationship between microbes where one consumes the waste products of another's metabolism (Pande and Kost 2017). Imagine that microbes A and B have a relationship where A consumes the waste-products of B, and B consumes the waste-products of A. If we attempt to grow A on a medium that contains all the essentials of A's diet except for the waste-products of B's metabolism, then A will not grow. Likewise for B. But, if we grow both A and B together on the same medium, both will grow as a community.

This kind of situation is a clear case of individual material allonomy and produces what Pande and Kost (2017) call "obligate metabolic interdependencies". W. F. Doolittle and I have elsewhere referred to this more generically as 'community interdependence' (Brunet and Doolittle 2018). Noting the existence of unculturability and community interdependence shows that microbes very often strike a balance between autonomy and allonomy. Since individual organisms within such communities cannot survive alone, cannot self-maintain, while a community of interdependent microbes can, it would be best to see the balance between autonomy and allonomy as an emergent community-level property (i.e. a property not possessed at the organism-level).

Explaining how these relationships evolved goes further towards showing that allonomy is a much deeper and widespread characteristic of life. There are two general explanations for the evolution of interdependence: it could be selectively favored, or arise neutrally. The first explanation has been championed by Morris et al. (2012) (and later by Fullmer et al. 2015) as the "Black Queen Hypothesis" (BQH). The latter by W. F. Doolittle and myself as the eponymously named "Gray Queen Hypothesis" (GQH) (Brunet and Doolittle 2018).[9] I will here review these two ideas, focusing on how they bear on autonomy and organismality, before showing that this sort of relationship applies to macrobes as well, ourselves included.

One reason that an organism might be dependent on its community is that it lacks a specific gene, and thus the capacity to synthesize an enzyme or complete a pathway—a form of mechanistic allonomy. The BQH and GQH differ in how they explain the loss of these capacities. What they both agree on is that, when an organisms lacks a specific self-directed capacity it can persist if another member of the community possesses this capacity in a way that is other-directed, or at least other-available. Morris et al. (2012) phrase this by saying

[9]The name is intended to contrast "neutrally" with the "black queen", but also after Michael Gray, who with P. Covello published the first paper deploying a similar sort of explanation (Covello and Gray 1993).

that a function (e.g. 'to degrade H_2O_2') is "leaky"; the function can benefit members of the community that do not possess the genes for it. Recalling the arguments of Ch.2, we might say that community interdependence can emerge when organisms already, fortuitously perhaps, possess extrinsically valuable effects. They explain,

> Gene loss can provide a selective advantage by conserving an organism's limited resources, provided the gene's function is dispensable. Many vital genetic functions are leaky, thereby unavoidably producing public goods that are available to the entire community. The BQH predicts that the loss of costly, leaky functions is selectively favored at the individual level and will proceed until the production of public goods is just sufficient to support the equilibrium community.—Morris et al. (2012) p.1

Allonomy due to gene loss can be selected for, under some circumstances. Indeed, according to Koonin and his collaborators, the loss and gain of capacities is the primary driving force behind microbial evolution (Koonin 2003; Koonin and Wolf 2012). Selection for conservation of resources by reduction of dispensable capacities is to be expected, and precisely this kind of evolutionary process leads to unculturability, community interdependence and both material and mechanistic allonomy. This is the selective or adaptive explanation for loss of self-directed capacities.

The availability of non-adaptive explanations shows that the phenomena are more likely.[10] Notice that once some organism B performs a leaky function (i.e. other-available capacity) any other member of its community A can lose (or reduce) their capacity to perform that function at no fitness cost. This also leads to interdependence, since now A cannot live in a community without B. That is, it need not initially be advantageous to lose a capacity, only non-disadvantageous, neutral. Add to this the fact that genes often mutate or are deleted at random during replication and the loss of genes and resulting community interdependence are likely outcomes even absent selection. This relaxation of the assumptions required to explain gene loss is the essence of the GQH. It is worth restating the distinction,

> [W]hile interdependencies can emerge from selection for genomic reduction, they can also emerge from neutral changes that are biased more towards interdependency than self-sufficiency (i.e. biased towards self-insufficiency)... At stake is whether genome reduction is selected for (as "streamlining") or is simply not selected against (as in Muller's ratchet).—Brunet and Doolittle (2018) p.18

[10] This is because assuming the conditions necessary for selection to take place restricts the cases that the explanation (BQH) can apply to, and because an explanation (GQH) not so restricted, applicable to a wider set of cases, is all-things-considered more likely to apply to an arbitrary case.

Put another way, the biological world is error prone but redundant across many selves. Self-directed capacities are constantly degrading or being removed entirely, but the presence of others that still retain those capacities can often be sufficient for the maintenance and organization of organisms. Selection may have a significant role to play, but mere neutral variation can result in the eventual change from self-maintenance to other-maintenance, or self-organization to other-organization. Biosynthesis does not occur in a vacuum separating self from other.

It is tempting to construe this as a peculiarity of microbial life—a realm already fraught with difficultly individualized entities (O'Malley 2014)—but analogously allonomous processes occur in macroscopic heterotrophic animals, ourselves included. Consider Payne and Loomis (2006) on the evolutionary loss of amino acid biosynthesis pathways,

> When an organism becomes a consumer by eating other organisms, all of the amino acids are available in the diet and no longer need to be synthesized. Unless amino acid biosynthetic pathways serve other essential functions besides providing amino acids, they are unnecessary and dispensable. Genes in dispensable pathways accumulate deleterious mutations, lose the ability to encode functional enzymes, and are eventually deleted from the genome. –Payne and Loomis (2006) p.275

Of the twenty amino acids required to assemble proteins only ten are universally conserved across animal life, with humans able to synthesize only eleven. Why these particular capacities could be lost is in part explained by the individual dependencies, or lack thereof, between capacities. Humans, as noted above, have retained their capacity to synthesize tyrosine because the capacity to degrade dietary phenylalanine depends on this. But these interdependencies are species and context specific. For instance, the parasite Cryptosporidium lacks both a tyrosine synthesis and phenylalanine degradation pathway, which Payne and Loomis (2006) hypothesize is facilitated by the fact that phenylalanine can be rapidly exchanged with its host. Here the loss of interdependent self-directed capacities and persistence despite allonomy seems more appropriately attributed to organizational openness, rather than closure.

Moreover, allonomy is not just a peculiarity of physiology, but also obtains in the formation of lineages and self-directed capacities of reproduction. In line with the epigraph from Maturana and Varela at the beginning of this chapter, when something is found to be capable of self-reproduction "the naïve approach is to deem it alive", and Mossio and Moreno explicitly take self-reproduction to be involved in "fully fledged" biological systems. Moreover, when an entity is observed to have a loss, reduction or transition in self-reproductive

capacity—such as in obligate endosymbionts—we tend to deny it other biological properties, such as autonomy and evolutionary individuality (Godfrey-Smith 2013; 2015).

The idea that variation in reproductive mode blurs the distinction between organismal and non-organismal biological systems has received much prior attention. Queller and Strassmann (2009) point out why we are primed for conceptual confusion about organismality: a number of closely related terms are given similar definitions grounded in intuitions about familiar organisms. Our use of 'unit of selection', 'vehicle', 'interactor' and 'individual' all seem to get us "unnecessarily tied up in issues that only philosophers love" [ibid p.3143]. Building on the work on evolutionary transitions in individuality of Maynard Smith and Szathmáry (1995), Queller and Strassmann (2009) begin with the idea that organismality hinges on a balance of cooperation vs. conflict. They then proceed with a survey—not unlike this chapter—of marginal cases of organismality. Importantly, they point out that cooperation and conflict are common, vary widely and often involve reproduction.

There are two ways to interpret Queller and Strassmann's (2009) survey of reproductive cooperation. Either we are equipped with a wide variety of marginal cases of self-reproduction in organisms, or a variety of marginal cases of organismality in self-reproducing systems. The first unquestionably supports the idea that organisms often lack self-directed capacities of reproduction; the latter is more problematic. If one is inclined to deny organismality to marginal self-reproducers, then it might seem as if we have vindicated component (3) in Mossio and Moreno's (2010) analysis of autonomy—non-self-reproducers (e.g. honeybee workers) would not be "fully fledged" organisms. This would be a misinterpretation of Queller and Strassmann. What Queller and Strassmann (2009) mean to show is that collections of organisms each individually engage in obligate forms of allonomous reproduction, and that this sometimes justifies *also* treating the collection as an organism. Organismality is not denied to lower level units when they form one at a higher level. A colony of honeybees is an organism, and so is a honeybee. Moreover, the cells of a honeybee are organisms on their view, each displaying a high degree of cooperation and low conflict, each obligated to reproduce cooperatively. In their words, "we have come to learn that all of our familiar organisms are superorganisms" [ibid].

Another account of variation in self-reproduction is offered by Godfrey-Smith (2013; 2015). Godfrey-Smith takes himself to be providing a general account of reproduction that can handle cases of symbiotic associations and transitions in individuality. I take him to provide a typology of recurring biological entities on the basis of varieties of mechanistic autonomy. Recurring structures can be first divided into reproducing and reconstructed. The latter includes biological structures like hearts and ribosomes, things that do not form

lineages and are recreated *de novo* by other things. Reproducing things can be further divided into *simple, scaffolded* and *collective*. He describes these as follows,

> A simple reproducer is something that can give rise to more objects of the same kind largely through the operation of resources internal to it—through its own biological machinery, in a broad sense—and further, is not made of smaller parts that also have this capacity [e.g. a bacterial cell]. A collective reproducer is a reproducing object that has parts that are themselves simple or collective reproducers [e.g. a multicellular organism]. Third, a scaffolded reproducer is an entity that reproduces (or is reproduced) in a way highly dependent on resources external to itself [e.g. viruses or genes]. –Godfrey-Smith (2015) p.70

The notion of collective reproduction allows the introduction of a hierarchy of potentially nested or cooperative reproducers, much like the reproductive allonomy described by Queller and Strassmann. The distinction between simple, scaffolded and reconstructed entities provides a different spectrum from autonomous to allonomously reproducing organisms[11]. A bacterial cell, perhaps only a non-symbiotic lithoautotroph, is a simple reproducer and is autonomous, it possesses every self-directed capacity of reproduction. But as an organism loses self-directed capacities and begins to depend on others it becomes a scaffolded reproducer, such as a symbiotic lithoautotroph or an obligate parasite. Finally, if something loses every capacity to self-reproduce and is entirely re-produced by another self then it is reconstructed (e.g., a heart, or honey-bee worker).

Intermediate forms of reproduction are widespread. The typology offered applies to classification of present diversity and of different stages in evolutionary transitions. Evolutionary history is full of simple-to-scaffolded transitions and marginal intermediates. Godfrey-Smith recounts the now standard view of eukaryogenesis: an alphaproteobacteria is engulfed by an archaeon, slowly loses reproductive capacities through selectively driven gene loss (BQH), and thus ends up a scaffolded reproducer within a collective one.

> Eukaryotic cells in different taxa today are at different places on a continuum from collective to simple reproduction, and their mitochondria are on a continuum from simple to scaffolded reproduction. –[ibid p.2]

The relatedness between the ancestral alphaproteobacteria and some modern intracellular parasites suggests that the initial stages of eukaryogenesis involved an ordinary parasitic,

[11] Godfrey-Smith (2015) indeed endorses a view of organisms much like that presented by Mossio and Moreno, where organisms are "metabolic units" and "systems that maintain their organization in the face of thermodynamic tendencies". He does not consider the possibility that an organism, in this sense, could be a reconstructed entity.

rather than fortuitous mutualist or physical "engulfing" interaction (Lynch and Marinov 2017; Brunet and Doolittle 2018), so the origin of eukaryotes may have resembled the allonomous situation of many present day organisms. Moreover, the transition from autonomous to allonomous reproduction need not involve obligate intracellularity. Viruses are paradigm cases of allonomous reproducers at present. However, one going hypothesis for the origin of viruses—which have intercellular as well as inter-species stages in their life-cycles—likewise involves their evolution from previously autonomous cellular parasites.[12]

The importance of these cases lies in the transitional stages between autonomous and allonomous reproduction. Along these evolutionary paths we find more or less familiar organisms, such as cellular parasites, that have lost specific capacities related to their self-reproduction. None are "fully fledged", since they must periodically return to the genomic or multicellular nest, but are perhaps "fledgling" autonomous biological systems. The absence of self-directed capacities often ties the reproduction of a "self" to one or many others, producing a situation where allonomy can prevail over autonomy. These cases simply cannot be understood as self-reproducing, autonomous, autopoietic organisms, yet they sit at the ecological and phylogenetic basis of all complex life. If autopoiesis is supposed to be the "organization which makes a living system a whole, autonomous, unity that is alive" (Maturana and Varela 1972), then it is at best a sufficient condition.

3.4.2 Machines That Possess Specific Self-Directed Capacities

The mechanical world is also diverse. The aim of the following sections is to show that machines are not totally allonomous or allopoietic, by examining some machines from the science of robotics. Both robotics and virtual machines have received considerable prior investigation as forms of "artificial life" (Bedau 2003); my intentions are different. For the present purpose, it is of no concern whether these machines are "alive" according to any of the myriad "definitions of life" that have been given (see Scharf et al. 2017; Mariscal and Doolittle 2017) except insofar as those definitions pertain to autonomous organization. Likewise for the origins of machines. It is equally irrelevant that these, and most, machines were created by humans and are largely the product of their intentions—if intention was the MOD (Ch.2) there would be no need to discuss autonomy. I concentrate on showing that these entities possess forms of autonomy in a literal sense.

[12]Transitions from allonomous to (more) autonomous reproduction are also possible: another hypothesis for the origin or viruses sees them as having "escaped" from early genomes, beginning as genomic parasites such as transposable elements. Likewise, there are cases of multicellular lineages giving rise to contagious unicellular cancer lineages in clams, dogs and Tasmanian devils (Forterre 2006; Metzger et al. 2015).

Autonomous capacities in machines cannot be comprehensively surveyed here (see Fong et al. 2003; Bedau 2003; Pfeifer et al. 2007). The field of contributions is wide, appearing in *Nature*, *Science*, and specialized journals like *Autonomous Robots*, *Robotics and Autonomous Systems*, various issues of *Artificial Intelligence*, and in the bulletins of private corporations and governmental bodies. My focus is on providing examples of autonomy in machines that contrast the forms of allonomy in organisms discussed above, under the headings of the broad self-directed capacities.

Self-Maintaining (Self-Repairing) Robotics—Self-maintenance requires the ability to regenerate some of the constraints (components) in a system while responding to external perturbations.[13] While simple forms of self-maintenance may be demonstrated by candle-flames or gas-motors, robotics has equipped us with more sophisticated capacities: self-healing materials, and self-repairing robotics. For instance, Markvicka et al. (2018) describe a 'elastic composite for soft robotics', consisting of a stretchable material with embedded liquid metal particles that "like natural skin or nervous tissue, are self-healing and remain functional even as material is torn, ruptured or removed". Hager et al. (2010) also describe some healing materials as autonomous when their response to damage does not require subsequent input from a human operator. A more mechanical form of self-maintenance or self-repair has been developed in the robotic arm of the International Space Station, named 'Dextre' and produced by the Canadian Space Agency, which is capable of "self-healing" by automatically replacing some of its parts.[14]

While being unable to regenerate specific components, some varieties of swarm robotics—robotic systems composed of many similar units—can excise damaged units and reform an approximation of their pre-damaged state. This type of self-maintenance was called self-assembly and self-repair by Tomita et al. (1999) (also see Rubenstein et al. 2014). Forms of self-maintenance that rely on a stockpile of available components to be replaced are somewhat dissatisfying, being a kind of mechanistic autonomy with material allonomy. But this same poverty of capacities is present in organisms, as in cases where absent amino-acid biosynthesis pathways are offset by amino-acid consumption. Dextre cannot persist without the availability of its components, but we cannot persist in an environment devoid of histidine. While distinct mechanistically from organismal self-maintenance, these are nonetheless cases where the existing constraints in a system help regenerate those same constraints and can respond to external perturbations.

[13]There are many examples of robotics that can maintain some behavior despite exogenous interference: Raibert et al. (2008) and Hutter et al. (2016) have both produced dog-like robots that can perform tasks despite being struck, falling or being pulled by rope. This kind of cybernetic or homeostatic adjustment to external stimulus is found in embryo in thermostats and any mechanical governor.

[14]Canadian Space Agency, January 2021, https://www.asc-csa.gc.ca/eng/iss/dextre/about.asp

What Mossio and Moreno add to the notion of self-maintenance is an emphasis on the role of energy flow. While a system may degrade from external damage, it might also exhaust its energy reserves. Various companies produce robots that accomplish phototaxis to overcome some energy loss. These are usually simple robots that can seek out sunlight and store the energy necessary for movement. For example, there are light-seeking houseplant pots, and robotic lawnmowers.[15] A more interesting case is provided by Iropoulos and colleagues at the Bristol Robotics Laboratory, who study "energetic autonomy" by linking robot behavior to chemical energy acquisition (Iropoulos et al. 2005; Melhuish et al. 2006). These authors have produced a robot called EcoBotII that can autonomously find sources of chemical energy (sugar cubes, flies or slugs), digest them in specialized Microbial Fuel Cells (MFCs), and store this energy for later use. Since microbes do most of the metabolic reactions required for energy production, they describe this as a variety of 'artificial symbiosis'. Here, the mechanical portion of EcoBotII is much like the giant tube worm described above: in both cases the resident microbes lose some energetic autonomy which is gained by the (robotic or animal) host.

Self-Organizing Robotics—Recall that the highest form of biological self-organization described by Mossio and Moreno requires that a set of constraints be collectively self-maintaining, i.e. that a set of mutually dependent constraints be closed. But also recall that this collectivity is partial: that only some of the constraints in a system are required to be in relationships of mutual dependence. This is a weakening of the view, highlighted above, that makes it possible to see many machines as engaged in—at least marginal—forms of autonomy. The self-maintenance of the robotic arm Dextre was used to replace cameras, if those cameras were required to measure or position arm movements, this would be a mutual dependence between two self-maintaining constraints (components). Likewise, the MFCs of EcoBotII require the constant input of energetic material to provide enough energy for motion in search of this material. Both Dextre and EcoBotII exhibit, at least marginal, forms of organizational closure between (some) pairs of their mechanical components.

A number of researchers in robotics claim to study both biological and technological "self-organizing systems"—sometimes referred to as "self-assembling systems"—drawing inspiration from the organization of tissues during cell differentiation or from swarms of social insects. R. Nagpal and colleagues offer a case with a peculiar magnitude of collective parts, KiloBot: 1000 small robots able to "self-assemble" into various shapes (Rubenstein et al. 2014). This kind of mutual dependence and collective interaction is not unique. The Institute for Electrical and Electronics Engineers has published volumes on "self-assembly"

[15] See article on houseplant robots by Amy McDermott (https://www.pnas.org/content/116/31/15313), and for robotic lawnmowers see Valify project (https://hackaday.io/project/72497-valify-v2-robot-lawnmower) (January 2021).

and "self-organization" since the early 2000s (see Yim et al. 2007). This self-organization and self-reorganization is collective, involves massive mutual dependence between parts, and partially regenerates some of the constraints that exist within a larger organization.

Self-Reproducing Robotics—The notion of a self-reproducing machine, automaton, began with the work of von Neumann in the early 1950s (von Neumann and Burks 1966), but was arguably not realized until 2003, when Suthakorn et al. (2003) produced "the world's first fully functional autonomously self-replicating robot". They built this machine out of LEGO. The robot is able to collect a number of parts, each corresponding to one of its subsystems, and assemble them into a whole that is able to perform the same task. Why robotics engineers would term this 'self-replication' instead of 'self-assembly' is evidently due to the fact that the assembly of a new machine is caused, not by the parts, but by another similar machine. Though this may be a form of self-replication, it is comparatively simple: there are only four subsystems, and the machine cannot assemble any of the parts of these subsystems themselves.

Some robotic systems surpass this and produce some of their parts. If we consider the individual units within a robotic swarm to be its parts, then the "self-replicating cubes" reported in Zykov et al. (2005) are a marginal form of self-reproduction or growth. The authors present robotic cubes that, when interlinked with a few copies, can build more such cubes from a number of parts provided—another case of mechanistic autonomy with material allonomy. More impressive is the actual production of parts used in copies—i.e. partial material autonomy as well. This was demonstrated in a proof of concept design by Adrian Bowyer and colleagues at the University of Bath (Jones et al. 2011; in dissertation, Sells 2009). The authors present what they call a 'kinematic assisted self-replicating and self-manufacturing machine', a 3d-printer that is able to print a "significant fraction" of its parts to be assembled by human operators. They call this machine RepRap.

The explanation of the genesis of RepRap is strikingly in line with my argument against an MOD based on autonomy. Jones et al. (2011) claim, as many in the field of robotics do, that their robot "was instigated by biomimetically considering extant naturally evolving strategies for reproduction". Interestingly, they do not locate inspiration for RepRap in an autonomous reproducer, but in the phenomena of symbiosis, noting, as above, that the lithotrophic microorganisms are the exception, not the rule, and that the vast majority of living things are interdependent. Jones et al. (2011) take this to imply that most reproducers, like RepRap, are 'assisted self-reproducers' (i.e. allonomous, scaffolded, or integrated reproducers). They argue that it is a proven evolutionary strategy to engage in reproductive mutualisms—citing Butler's *Erewhon*—similar to that between bees and flowers. The machine provides some additional parts useful to us in exchange for our symbiotic assistance in copying. The first

machine was completed in 2008, and in 2011 there were 4500 across the world stemming from this 'ancestor'—an apparently successful mutualism for RepRap.

3.5 Conclusion: Spectra of Autonomy

All of the autonomous robotics evidenced today are semi-autonomous, none reach the level of "fully fledged" autonomy. Neither do many organisms. The autonomy-allonomy distinction spans and cuts across biological and technological diversity. Maturana and Varela were correct when they said the "naïve approach" was to deem something alive if autonomous—there are *prima facie* non-living things with some autonomy. And on the contrary, there are living things with a significant degree of allonomy. When Mossio and Moreno talk of increasing autonomous capacities of self-maintenance and self-organization or closure leading to a "crucial transition between the physiochemical and the biological domain" or "a fundamental property of... biological organisms", this is only half of the picture—the other half involving the "technological domain" or "automatons". Autonomous machines likewise possess forms of these capacities as "crucial transitions" or "fundamental properties", distinguishing them from their familiar relatives.

In this chapter I have addressed the diversity of autonomy and its distribution in machines. The Kantian grounding of teleology (for Nicholson 2013) or autonomy (for Maturana and Varela 1972; and Moreno and Mossio 2010) in reciprocal causation among parts, in features of "organization", is only applicable at a very coarse grain of analysis when applied to a narrow sample of living and mechanical things. When we look at more fine grains of analysis, to the refined self-directed capacities, and to a broader range of organisms and machines, the gap between autopoietic and allopoietic, between autonomous and allonomous, begins to be filled. I take this to show that an MOD cannot be made on the basis of autonomy.

But neo-Kantian accounts of organization are not the only way in which the internal makeup of organisms has been used to distinguish them from machines. In the following two chapters I discuss two other contemporary views of the MOD that trace back further than Kant. First (Ch.4), Dupré (2018; 2017; 2013; 2012) and Dupré and Nicholson (2018) have shifted the focus from the merely reciprocal nature of causation in organisms to its *processual* nature, tracing this view back, much farther, to Heraclitus. Next (Ch.5), a cluster of views falling under the umbrella of neo- or New Mechanism (Machamer et al. 2000; Bechtel 2011), has also been argued to provide a metaphysics of biological causation, itself dating back to older mechanistic ideas, most significantly to the mechanical philosophy of Descartes. Perhaps, one may hope, Descartes or Heraclitus can provide the MOD that Kant could not.

The following two chapters argue that they do not provide the MOD and, moreover, that the metaphysics of process and mechanism are better off without it.

Chapter 4

Process Ontology and the Machine Organism Distinction

Abstract

I argue against MODs lodged within process ontology, represented primarily by the views of John Dupré and Dan Nicholson. Process ontology has encouraged philosophers to think differently about organisms, though it has done so at times by advancing MODs. These MODs fall generally under the headings of "constitutivity of processes", "material flux", and "hierarchical stabilization". I show that there are machines today that enjoy the properties that Dupré and collaborators advocate as specifically organismal. I take this *not* as a rejection of process ontology, but as an argument against the MODs thought to support it, and as an argument for a processual ontology of technology insofar as a processual ontology of biology is appropriate.[1]

Introduction

This chapter is much like Ch.2-3: I will consider another group of MODs on offer in the literature and argue against them. It is also different in two ways. Firstly, this and the next chapter are the first to deal with the MOD as a problem of metaphysics. Secondly, my aim in this chapter is more synthetic and less negative than Ch.2-3.

Process ontology (AKA 'processualism', 'process metaphysics', or simply 'process philosophy') and neo-mechanism (Machamer et al. 2000) are often positioned as as rival metaphysics—each competing to provide the correct account of organisms, and each dealing

[1] This work was presented at the 2019 meeting of ISHPSSB, Oslo.

differently with how to distinguish organisms from machines.[2] A more in-depth discussion of new-mechanism will follow (Ch.5), though it will sometimes be necessary to discuss and contrast both here. The hope for processualism is that it can provide us with a metaphysical account of organisms. In recent work by Dupré and collaborators, justifications for processualism have relied on MODs that are subject to naturalistic counterexamples. I see the provision of MODs as an improper justification for process ontology, and argue against them here. However, the merits of process ontology are much broader than the specific issue of the MOD. Instead of pushing back against a processual-MOD by arguing against process ontology, and for neo-mechanism, I argue that machines have a place within process ontology. The overall aim of this chapter is to argue that being properly subjected to a processual analysis is not an MOD.

This chapter consists of four main parts: In § 1 I provide an overview of process philosophy of biology, arguing that it is a viable metaphysical account of biology, despite some problems. In § 2 I explain how the metaphysics of process philosophy is connected to the MOD. In § 3 I detail some of the MODs offered within process ontology, focusing on those of Dupré and Nicholson (2018), and argue that these MODs are neither true nor supportive of process ontology. Finally in § 4 I argue that, far from needing to exclude machines from a process ontology of biology, we are better off by integrating machines into a "process ontology of technology", and I provide some elaboration of what that should look like.

4.1 Process Ontology is a Contender

Modern proponents of process ontology in science make use of concepts of process originating from ancient, early-modern and modern sources. I begin with a short philosophical history of processualisms leading up to its modern formulation within the philosophy of biology. What is perhaps new in modern process philosophy is its application as an analytic philosophy of science. The role of process philosophy in the philosophy of biology in particular has been called a "trend", "rise", "turn", "paradigm shift", "revolution", "reconceptualization" and even "revisualization" (Seibt 2018; Dupré and Nicholson 2018). It has received substantial attention in the literature and its advocates have replied to a variety of criticisms (see § 4.2). It finds one of its most prominent defenders today in John Dupré, who has become the modern champion of a specifically biological process ontology (Dupré 2018;

[2]Since my issue is only with the latter, I will not directly address the disputes between these philosophies except insofar as it bears on the provision of an MOD by either.

2017; 2013; 2012; Dupré and Nicholson 2018; Dupré and Bapteste 2013; O'Malley and Dupré 2005; also see Woese 2004).

Outside the philosophy of biology, general process philosophy is often characterized as a rising underdog. This is not quite right: process philosophies broadly speaking are coeval with the origins of philosophy, and processual concepts coexist with substance and mechanistic ones throughout scientific and non-scientific discourse. Processualism of some form has been around since antiquity. Rescher (1996) traces process philosophies from the 'flux' of Heraclitus in the 5th century BCE, through Plato and Aristotle, Leibnitz, Hegel, Peirce, James, Bergson, Dewey and Whitehead. It is common to attribute the origin of process philosophy to Heraclitus. The attribution is based on a single fragment of his writing, the so-called river passage, which is variously translated and perhaps even originally ungrammatical.

> On those who enter the same rivers, ever different waters flow. —Heraclitus (c. 470 BCE)

The river passage, though *prima facie* an innocuous comment about the non-identity of successive portions of moving fluid, is interpreted by process philosophers as offering the first profound claim about the identity of certain kinds: some things stay "the same" by constantly changing.[3] Though an ancient source for contemporary process philosophy, it is dubious whether Heraclitus should mark the origin of process philosophy. For one, it is plausible that Heraclitus was not advocating an idiosyncratic metaphysical view of change at all, but was instead merely providing another example of his broader philosophy of the "unity of opposites". In connection with other fragments of Heraclitus, e.g. "The way up and the way down is one and the same" (Burnet 1908, fr.69), the river passage can be seen as just another attempt to unify the opposites of "sameness" and "difference". And there are other ancient sources, less mired in interpretive difficulties.

Even before Heraclitus, in India (c. 480 BCE), the last words of the historical Buddha are sometimes reported to begin as "Conditioned[4] things are by nature perishable…". And later, in Japan (c. 1210 CE), the Buddhist monk Chomei left notes with a passage nearly identical to Heraclitus's river passage, almost surely of independent origin.

> On flows the river ceaselessly, nor does its water ever stay the same. —Chomei, c. 1210 CE

[3]The above translation is, perhaps, a more accurate rendering of the river quote: "One cannot step into the same river twice." This is not found in the fragments of Heraclitus. The river quote seems to be preferred by some since it makes a stronger claim about the impossibility of identity over time for changing entities.

[4]There is a vast literature on the correct interpretation and translation of 'Sankharas', above rendered as 'conditioned things'.

These early philosophies leave us with processualism in an innocent form: change is emphasized over stasis. Later versions of process philosophy elevate it to a metaphysics in its own right and almost invariably position themselves in opposition to another metaphysics. That rival, depending on the advocate, is either essentialist (substance) metaphysics, mechanical philosophy, or neo-mechanism—Locke, Descartes, or Machamer et al. (2000)[5].

The next major leap in process philosophy came with the speculative metaphysics of Alfred North Whitehead's (1929) *Process and Reality*. Though largely impenetrable,[6] Whitehead's philosophy does contain perhaps the first modern and general formulation of a tenet of process philosophy that was—in Heraclitus and Chomei at least—only gestured at.

> How an actual entity becomes constitutes what that actual entity is... Its 'being' is constituted by its 'becoming'. This is the 'principle of process'.—Whitehead (1929) p.31

Whitehead's formulation of this principle of process is a more explicitly processual version of the earlier statements of Heraclitus, Chomei and the Buddha.[7] Whitehead termed his view the 'philosophy of organism'. Perhaps processualism as a philosophy of biology begins here, although his processual 'philosophy of organism' was hardly what we would think of today as a process-based account of organismality. His view instead presumed a kind of processual nature to organisms and provided a re-description of even non-biological things—the cosmos, god—in organismal terms. His view was panpsychist and organicist, attributing mind- and organism-like properties to non-minds and non-organisms.[8]

In a sense, process ontology today is a neo-processualism. It has abandoned Whiteheadian (and Bergonsian) metaphysical terminology, and largely—though not completely—separated itself from organicism and panpsychism.[9] Although the move from the mechanical philosophy of Descartes to neo-mechanism was an almost total metaphysical refurbishment, the

[5] More on mechanism below, in § 4 and Ch.5

[6] Apparently in all seriousness, Dupré (2013) remarks on the worry that 'process' is poorly characterized that "it is widely feared that taking processes seriously as a basic ontological category will require reading Whitehead, who is rumored to be unreadable and possibly unintelligible", but maintains that that biological pay-offs of the notion are so great that "we have a real motivation to think seriously about process—even if it does mean reading Whitehead" [ibid, p. 31].

[7] Even this comparatively succinct statement in Whitehead requires some disambiguation and unpacking. The reference to 'actual entities' could easily be misinterpreted as a concession to mechanism, especially since neo-mechanists make 'entities' part of their dualist ontology. Whitehead is rife with synonyms for 'a specific process', which are indeed sometimes called 'entities'. Likewise, though 'becoming' certainly includes the more narrow notion of "the origin of something" it should not be restricted to that use. Whitehead seems to treat 'becoming' as a general word for what might be better, and sometimes is, rendered as 'processes of change'.

[8] It was 'organicist' only in retrospect, since 'organicism' proper came after Whitehead.

[9] Though Dupré and Nicholson (2018, p.11, fn.8) are careful to distance themselves from parallel projects that have attempted to flesh out a process based philosophy of biology from Whitehead's more panpsychist comments, Dupré (2017) concludes with a "speculative" thought that "agency is a much more natural attribution" to processes, rather than things or substances. Denis Walsh (Dupré and Nicholson 2018, p.167) comes to similar

change to process ontology of biology from earlier process philosophies is more continuous—as is fitting. Dupré (2017) makes use of Rescher's (2004) general process philosophy, and Dupré and Nicholson (2018) acknowledge some antecedent roots (of inspiration at least) in both Whitehead and the organicists.

> [I]t is perhaps more appropriate to describe organicism as a philosophy of biology that was inspired by Whitehead than as a genuinely Whiteheadian philosophy of biology... We regard the organicists as kindred spirits and consider our project to be continuous with the earlier tradition in the philosophy of biology to which they belong.—Dupré and Nicholson (2018) p.11

There are other process ontological views—process theists and various process philosophies of physics—but in Dupré and collaborators we find a process ontology of biology. The following section dissects that philosophy into four features that are thought to apply specifically to organisms, to prepare for the arguments § 4.4 that indeed these features are not specific to organisms but apply to some machines also.

4.2 Four Key Features of Process Ontology Today

In this section I will spell out what I take to be the main features of processualism, with attention specifically on its consequences for our analysis of organisms.

4.2.1 Constitutivity of Process

Aside from metaphysical precursors, Dupré (2012) pins some antecedents of his view in the Developmental Systems Theory (DST) of Griffiths[10] and Gray (1994).

> DST teaches—or reminds—us that there is much more than the passage of genes needed to transmit living form from one generation to the next... [DST] points us towards a fundamental fact about organisms: they are *not properly understandable* in terms of one set of properties, say those of the adult organism, but are ultimately processes.—Dupré (2012) p.2, my emphasis

Clearly, one can "understand" a great deal about organisms by simply looking at the static characteristics of adult specimens—perhaps by freezing them and taking cross sections,

conclusions under his heading of "methodological vitalism". While certainly not panpsychist, this kind of description of processes by analogy with agents is a sort of "panagentialism".

[10]Griffiths and Stotz (2018) furthermore trace their emphasis on developmental systems and de-emphasis of adult forms to Waddington (1952).

or basing theories on photographs taken in the wild. Moreover, there are plenty of scientific contexts where precisely this kind of understanding of organisms would be "proper"—such as in anatomy, or demography. I take Dupré's point to be about a deeper, more fundamental understanding. When Dupré (2012) says that organisms are not "properly understandable" in terms of one set of (adult) properties, it is best to interpret him as offering a claim about how processes of change are constitutive of being an organism. Perhaps, in a slogan: *ontogeny is organismality*.

This is plausibly a biological instance of Whitehead's "principle of process", that an organism's being is constituted by its becoming. That it is (often cyclic) living processes and not particular, static, sets of properties that are the proper objects of biological inquiry is perhaps the most fundamental claim of process ontology of biology.

> [T]he organism encompasses the entire life cycle; indeed, it is the life cycle itself that constitutes the organism. —Dupré and Nicholson (2018) p.19

When one studies static sets of an organism's properties, one just isn't studying everything that organism encompasses. For example, when one studies the anatomy of adult graylag geese, one is explicitly not studying graylag geese *in toto*. An organism in perpetual stasis ceases to be the organism that it was (in a sense to be explicated in § 4.3.1). When a hibernating frog is caught in glacial ice, instead of temporarily frozen pond mud, it is no longer going through the seasonal changes characteristics of its life cycle. It is not just the properties of the adult that count. The changes in development (and degeneration) are fundamental aspects of what it means to be an organism.

We apparently cannot escape this DST/processualist conclusion in molecular biology either. That we cannot properly understand an organism in terms of one set of properties holds for genetic properties too, according to Dupré (2011; 2017; 2018). While a gene-essentialist approach to studying organisms with changing life cycles might identify them with the constant properties of the genome, Dupré (2011) argues that it cannot serve this role, since the genome also changes with development. Dupré (2017) summarizes this position by pointing out that,

> [G]enome sequence is as much the consequence of organismic stability as it is its source. For a process... no such constant property is required.—Dupré (2017) p.2

A processual view of organisms does not see them as "things", constant or persisting sets of (essential) properties, either at the level of gross phenotypic nor molecular and genomic traits.

4.2.2 Hierarchical Stabilization

While process ontologies are resistant to seeing organisms or life cycles as "things", substances, or entities, certain processes nonetheless enjoy varieties of individuality, holism, temporal extension, persistence, and re-identifiability. To the process ontologists, we simply do not need things or substances to support these latter, important, concepts. For instance, that there are no constant characteristics shared by different stages of a life cycle does not prevent us from considering these stages as parts of a whole or "individual process" (Dupré 2017).[11] Process ontology of biology nonetheless does differ in how individual processes are explained.[12] For things, the major problem is explaining how and why they change. For processes, change is a given, so we need an account of their stasis or "stability". Process ontology of biology demands an explanation of the stabilization of organisms as individual processes.

At the level of general process philosophy, the outline of an explanation for the stability of an individual process is quite clear. Since all processes are thought to be "dynamic" or in "flux", the explanation for the apparent stability of a given process will need to be provided in terms of particulars about the dynamicity or flux of other processes. Stable processes are thought of as a confluence of (near) equal and (near) opposite effects of other processes, themselves potentially unstable. Dupré and others moreover claim that only relative stability is available; stability is only ever apparent over a given timescale. Some processes will be stable, entity-like, though this is an abstraction justified only in cases of scientific explanation where that stability is constant over the timescale of the explanation (Dupré 2017). All apparently stable processes are, over a longer timescale, unstable. Explanations of the stability of particular processes over a given scale then depend on their treatment within a specific science (§ 4.5).

Consider a vortex. A vortex forms when fluid moves in a rotating path around some axis. In a whirlpool or tornado the axis meets the center point; in a smoke-ring or mushroom-cloud that axis is curved into a circle. A vortex is certainly something we individuate (more or less precisely) as a pattern of flux.[13] Unlike comparatively more static entities of the same overall shape—a single candelabra or a metal ring—explaining a vortex requires an account of how it stabilizes in that particular configuration (during a given time period). This explanation will differ depending on the vortex, but will invariably require reference to both the internal processes within the vortex (such as the rotation of air) and to external processes in its

[11] Seibt 2018) works towards defining a "processual mereology" along similar lines to the traditional mereology of substances.

[12] This is Dupré's terminology, inspired by the use of "individuals" as opposed to "classes" in Hull and Ghiselin.

[13] The term 'flux' in its mathematical sense is defined via motion through a surface or body.

environment (such as the formation of temperature gradients in the atmosphere). It may also require reference to processes at different temporal and physical scales, such as the molecular characteristics of the fluctuating substance or seasonal variation in atmospheric conditions.

In the specifically biological context, this idea that processes require explanations for their stability has been unified with an account of the biological hierarchy to provide an account of "hierarchical stabilization".

> [A]n organism is not organized as a hierarchy of structures (as a machine is), but as a hierarchy of processes... Accordingly, the stability of a multicellular organism as a whole derives from the continuous regeneration of its tissues, which are themselves maintained by the incessant renewal of their cells, which are in turn stabilized by the ongoing replenishment of their molecular constituents.—Dupré and Nicholson (2018) p.16

Organisms conceived as processes are stabilized over diverse timescales at different hierarchical levels. During its life cycle an organism is stable, homeostatic, in the face of changing internal and external conditions, such as metabolic activity and environmental changes. Moreover, species or lineages may remain stable over evolutionary timescales—consider the relative evolutionary stability of *Crocodylus*. Dupré (2018; 2017) and Dupré and Nicholson's (2018) *Manifesto For A Processual Philosophy of Biology*, address two important aspects of this view, one below and one above the level of organisms. These will be addressed in turn before turning to the relationship between the MOD and process ontology (§ 4.4)

4.2.3 Material Flux

Process ontologists consider the flow of matter during metabolism to be both evidence for their view and, more strongly, correctly understood within it. Dupré and Nicholson (2018, p.15) say that metabolic turnover is "one of the strongest motivations for adopting a process ontology" and that "from a metabolic perspective, it is simply a matter of fact that, in an organism, everything flows" (ibid, p.17). The guiding idea is that some aspect of the flow of matter and energy through the bodies of organisms, such as its metabolic or thermodynamic nature, suggests a processual over a substance, thing-based or machine-based ontology. Notice that this simply follows from a rejection of a thing-based ontology if we consider machines to be paradigm things, and not processes. Nicholson (2018) at times says as much: "After all, what are machines if not persistent material things with determinate sets of

properties…" (in Dupré and Nicholson 2018 p.141). Carl Woese,[14] a founder of molecular biology, points to precisely the same thing.

> Machines are not made of parts that continually turn over, renew. The organism is. Machines are stable and accurate because they are designed and built to be so. The stability of an organism lies in resilience, the homeostatic capacity to re-establish itself.—Carl Woese (2004) p.176

Process ontologists cite the often surprising rates that biological structures are replaced or repaired. Moreover, processualists note how these rates only seem to increase as we move down the (organizational) hierarchy, from tissues to cells and finally macromolecules. Dupré and Nicholson (2018, p.17), citing the New York Times article *Your Body Is Younger Than You Think* by Nicholas Wade (2005), point to the fact that the liver as a whole is replaced annually, the entire skeleton is replaced on average every 10 years, and the protein turnover rate is approximately 8% daily.[15]

The significance of a processual understanding of metabolic turnover is easily underestimated. Non-processualists plainly do not deny the events of turnover and often explain it mechanistically, so it is hard to see why turnover would support an alternative metaphysic.[16] To process ontologists, the importance of metabolic turnover is not simply a shift in emphasis towards the replacement or renewal of parts over their constancy or persistence—the persistence of parts is precisely what the process ontologsts believes is most in need of explanation. Instead, it can be taken as a call for metaphysical reform in analyzing claims about physiology, metabolism and cell biology (see Nicholson 2019). Indeed, if the flow of consumed and metabolized material is supportive of process ontology, then some claims about flux within modern molecular biology seem deeply confused.

The term 'metabolic flux' is used in molecular biology to refer—sometimes equivocally—to both the rate and actual event of biochemical transformation or movement of metabolites, either between states or between compartments. For example, the (rate of) transformation of fructose-1,6-bisphospate (FBP) to fructose-6-phosphate (F6P) during gluconeogenesis by the enzyme Fructose-1,6-Bisphosphatase (FBPase) is the metabolic flux of FBP through FBPase to F6P. Kochanowski et al. (2013) have shown that this reaction is governed by a "flux sensor", i.e., that the rates of production and overall concentration of FBP reciprocally affect

[14] Woese's evolutionary theory is addressed in Ch.6.

[15] These results indeed require more cautionary interpretation than Dupré and Nicholson provide. For instance, a 8% /d protein turnover rate might mislead that *all* proteins in the body are turned over on average every 12 days, but that would require more metabolic activity than any human body is capable of. I will discuss these and similar results critically below.

[16] I take it that "events of change" is metaphysically ambiguous between "mechanism of change" and "process of change", which I intend here to avoid begging any questions against either side of that debate.

the production of the protein FBPase. Consider the admixture of computational, cybernetic, metabolic, mechanistic, and processual concepts in their discussion,

> Flux-sensing mechanisms can be integrated into regulatory control circuits: a nutrient flows into a cell and realizes a metabolic flux that is being measured and used for regulation of protein expression, which (in the case of enzymes) will in turn influence flux, overall resulting in robust control loops.—Kochanowski et al. (2013, p.4)

If Dupré and others are correct that material flux through organisms is a strong motivation for adopting a process ontology, then mechanistic claims like that of Kochanowski et al. (2013) above are, at least, suspect targets for reform. A (neo)-mechanist explanation of metabolic flux in terms of persisting entities will apparently be incorrect if any constituent entity is dynamic. According to Dupré this poses a problem if the entities are dynamic over the timescale of the events explained. Likewise, Nicholson (2019) notes the non-mechanistic nature of gene regulation of the sort needed to regulate FBP flux. In this case of regulation of flux, and similar, the dynamic changes in an enzyme is an event to be explained: enzyme concentrations are dynamic over the course of FBP metabolic flux. On the processual perspective, the phrase 'the mechanism of a metabolic flux sensor' is a misnomer: in reality, there are only lower level metabolic processes that stabilize higher level ones, such as *the process of FBPase inhibition stabilizing the metabolic flux through gluconeogenesis.*

4.2.4 Top-Down Evolutionary Stabilization

> [I]t is self-evident that evolution is a process, it is less clear how we should think about the nature of the entities that participate in this process.—Dupré and Nicholson (2018) p.34

I have up to this point considered hierarchical stabilization of individual processes with respect to external environmental processes and internal metabolic ones. What remains is the processual account of stabilization at levels higher than the organism: in species as evolving lineages. It is obvious from the forgoing that the evolving entities will turn out to be processes: some developmental systems or life cycles are sufficiently integrated to form lineages, which are temporally extended processes that are the participants in evolutionary change (Dupré 2018).[17] What is interesting about including the reference to evolution as a process is seeing

[17]Dupré is somewhat loose as to which life cycles are sufficient—in Dupré (2017) he argues that only sexual species are individuals, whereas Dupré and Nicholson (2018) seem to follow (and expand on) the more traditional line of the Hull-Ghiselin thesis in arguing that all species are individual processes, including perhaps asexual ones. Although, this fits well within the scheme of Dupré's overall commitment to promiscuous realism.

evolution as imparting stabilization on lower levels of the biological hierarchy.[18] Essentially, process ontologists note the processual nature of stabilizing selection.

Trans-generational or lineage stabilization is in part affected by biochemical processes that ensure reproductive fidelity and can repair some forms of genetic alterations. The process ontologist points out that this bottom-up explanation of lineage stability is insufficient. DNA repair mechanisms are imperfect, and reproductive fidelity is too limited to produce the kind of stabilization enjoyed by lineages over long periods of time. Think crocodiles again. Additionally, we must account for the top-down effects of stabilizing selection, i.e. the death or reduced reproduction of un- or less-fit variants taking place continually as they arise. Indeed, lower-level biochemical processes also introduce variation, most of which is deleterious and destabilizing for lineages. This must be selected away "at a higher level" to maintain stability. Moreover, for asexual species at least, we must expect that stabilizing selection contributes significantly to the explanation of stabilization, sufficiently at least to oppose the constant introduction of mutations—stabilizing selection must be able to wrench Muller's ratchet.

In the above I have uncritically stated what I take to be the major points of process ontology (as advocated by Dupré and collaborators). We are given a monist metaphysics where the fundamental ontological category consists of processes, which are seen by processualists as providing the identity or individuality conditions for familiar entities (organisms, lineages) and explained using a hierarchical framework that focuses on their stabilization, as opposed to stasis, in the face of otherwise constant flux. In the following section (§ 4.3) I will show how process metaphysics has been connected with the specific project of defining and defending an MOD. Essentially, process ontology together with the assumption that there is some MOD leads to the conclusion that it is within a process ontology of biology that such an MOD should be made. In § 4.4 I will examine specific MODs offered by process ontologists, arguing both that they fail and that, even were they true, they do not speak in favour of process philosophy.

4.3 Relationship Between Process Ontology and the MOD

In this section I argue that the provision of an MOD by process philosophers amounts to a specific and independent processualist theory under the broader heading of process ontology

[18] This seems to be "self-evident" when we consider the alternative "evolution is a thing", a patent category mistake. But when we consider the alternative that "evolution is a mechanism" there certainly are biological discussions of "mechanisms of evolution" (see Rose and Doolittle's (1983) *Molecular biological mechanisms of speciation*).

of science. Thus, we can argue against processual MODs without denying process ontology in general. To begin, we should separate out three theses of process philosophy (MT, ST and DT below), the latter two of which relate directly to the philosophy of biology. The distinction thesis (DT), last below, is connected to the MOD and will be my primary concern.

MT Metaphysical Thesis: The world is ontologically composed of processes, i.e. not of mechanisms, things or substances.

ST Scientific Thesis: The correct ontology of biology is a process ontology, i.e., ontological questions about biology are successfully settled by processualism.

DT Distinction Thesis: Process ontology successfully distinguishes biological entities from mechanical ones. Process ontology of biology successfully analyses the distinction between machines and organisms.

Each thesis has defenders. Whitehead's philosophy of organism and Bergon's Creative Evolution are both metaphysical and scientific theses. Dupré (2018; 2017; 2013) often defends both MT and ST. Although with coauthors, such as Dupré and Bapteste (2013) and O'Malley and Dupré (2005), Dupré has focused more directly on processual varieties of specific scientific theories in biology—microbiology and systems biology, respectively. Since (neo-)mechanism is the de facto rival scientific ontology, many authors offer something in support of ST (e.g. Dupré 2018; 2013; Dupré and Nicholson 2018, c.f. Austin 2016). I will concentrate on the versions of the MOD and DT advocated in Dupré and Nicholson (2018).

The ordering of the above theses is important. If we assume that machines and organisms are really ontologically distinct (i.e. MOD, not DT itself), then justification runs up the list while implication runs down: DT justifies ST, which in turn justifies MT, while MT implies ST, and assuming the MOD, ST implies DT.

4.3.1 Justification of Theses

The distinction thesis is used as justification for the scientific, and the scientific for the metaphysical—although none *implies* the next in this order. In the first case, if process philosophy can settle the issue of the distinction between machines and organisms, this give us reason to think that process philosophy provides the correct ontology of biology. Were it the correct ontology of biology, then it should settle any real distinctions between elements of that ontology (organisms) and others (machines). This seems to be the implicit reason for many to address machines in advocating for a process ontology of biology (see Austin 2016, and § 4), but few are more explicit about it than Nicholson (2018).

> Demonstrating the ontological inadequacy of the MCO [machine conception of the organism] is a necessary first step if we are to come to terms with the processual nature of life and lay the foundations for a processual philosophy of biology.—Nicholson (2018, in Dupré and Nicholson 2018, p.141)

This kind of justification is abductive in the sense that it justifies on the basis of (hypothetical) best explanations. It is not inductive, in a more probabilistic sense: the correct ontology for theories is certainly not often or always determined on the basis of its ability to distinguish its objects from machines. Instead, we are invited to think that the correct ontology of biology should be one that is at least capable of distinguishing machines from organisms. Thus, if processes ontology can do that, it satisfies at least this necessary condition.

Secondly, the scientific success of process ontology of biology provides a reason that it is the correct metaphysical account of reality. Dupré and others also point to the apparent success of processual theories in others sciences, such as physics, for truly inductive support. This latter justification for process metaphysics is less contentious. Surely, if we intend to be naturalistic about reasons for theory choice in metaphysics, then we ought to look to the ontologies actually or ideally used by natural sciences. If biology and many other sciences are best analyzed with an ontology of process (ST), then a good (naturalistically: the best) explanation for this is that every domain of the world is processual (MT). In the final section (§ 4.5) I in effect bolster this claim myself by arguing that, insofar as we accept a process ontology of biology, we have reason to accept a process ontology of technology—further reason to favour the overall metaphysical thesis.

4.3.2 Implications Between Theses

Arguing that, given process metaphysics, process ontology of biology ought to establish the MOD is easier. The central tenet of process metaphysics is that processes are the fundamental ontological category; the MOD is, in general, simply the claim that machines and organisms are ontologically distinct. Together these imply that the MOD will be, fundamentally, about processes. This can be cashed out more fully in terms of the three theses above.

Firstly, if the world is ontologically composed of processes (MT) and not things, then the correct account of any particular domain of the world will require an ontology of process. So (ST) biology, being a theory about a particular domain of the world, requires a process ontology. If everything is a process, biological things are. Secondly, if machines and organisms are in fact ontologically distinct (MOD) and process philosophy is the correct ontology of biology, then (DT) it is within process ontology that the ontological distinction between machines and organisms will be settled (presuming it is settled at all). Assuming MT

and the MOD, this is straightforward. Nonetheless, note that we need to assume the MOD to move from ST to DT. That is because the correct ontology cannot distinguish what is not ontologically distinct, under pain of inconsistency. Simply claiming that the correct ontology of organisms is processual does not, by itself, imply that machines are non-processes. This is often overlooked when opposing machines to organisms on processual grounds. In § 4 I will argue that indeed DT fails. For the moment, consider what this would mean for process ontology of biology.

Consider any specific theory under process ontology of biology, dealing with a specific sort of biological phenomena. Take Barwich (2014; in Dupré and Nicholson 2018), who argues for a processual account of olfaction. Though we may think that process ontology must provide some theory of olfaction if it is indeed the correct metaphysics of biology generally, we would also not want the truth of MT to hang on the correctness of that particular account. This is reminiscent of the debate between Harvey and Descartes on the correct account of the movement of blood though the heart (Gorham 1994). Although strange today, Descartes believed that certain tenets of his metaphysical system turned on details of his theory of the pumping of blood—which involved the heating of blood "drop by drop". Specifically, it seemed to Descartes that the alternative implied vitalism and denied his mind-body dualism. While perhaps laudable as naturalistic metaphysics, this tying of metaphysical to physiological details is unnecessarily foundational. Descartes simply did not recognize that his theory of the motion of blood was just one of many possible Mechanistic explanations, not a keystone issue. Likewise, we would not want the truth of MT to hang on the correctness of Barwich's processual account of olfaction, since it is one of many potential processual accounts of smell. The ways that process ontologists attempt to distinguish machines from organisms are not integral parts of process philosophy of biology, but rather specific processual theories under that umbrella.

This fits well with the arguments of previous chapters: if we tie our establishment of an MOD to scientific theses about the ontology of biology, then we unnecessarily make our biological ontology dependent on (largely) historically contingent facts about the present diversity of mechanical forms—in this case, their diversity in comparatively processual features. This is a mistake we can avoid, on the one hand (§ 4.4) by keeping the MOD and biological ontology separate, and on the other (§ 4.5) by showing that any sufficiently general processual scientific ontology can apply to technology as well.[19]

[19] In § 4.5 I define what I take to be two necessary criteria for general process ontologies of science.

4.4 Against Processual MODs

> Organisms, due to their thermodynamic nature, are metabolizing entities that must continuously act to continue to exist, machines can do nothing (for lengths of time) without ceasing to exist. While you can leave your typewriter unattended for a decade, you cannot do so with your hamster.—Dupré and Nicholson (2018) p.15

There is an internal tension in the recent process ontological turn in biology. If process ontology of biology is to secure a special place as a metaphysic of biology, then it seems as though it must displace or supersede alternative approaches, one of which being the conception of organisms as machines (see Nicholson 2013; Nicholson and Dupré 2018; c.f. Machamer et al. 2000, who argue that processualism and substantivalism are unified by new mechanism). But the broader metaphysical project of process ontology would be furthered by providing a processual account of machines as well. I begin with this general tension before turning to specific processual MODs.

Confronted with the problem of providing an MOD that justifies process ontology *of biology*, one could of course deny *panta rhei* ('everything flows')—that is an analytic option, but one not taken by processualists for good reason. One can see why: if process ontology of biology rested on a more restricted *panta (biological) rhei* then it would need to define the conditions for being a biological thing, an unhappy demarcation problem to solve before the larger efforts of process ontology of biology could begin. On the other hand, even equipped with demarcation – such as teleology (Ch.2) or autopoiesis (Ch.3) – it would need to defend the odd claim that biology demanded a special fundamental ontology.

Much to the contrary, in order to gain traction for a process ontology, Dupré and Nicholson (2018) and others look to the success of a process view in other sciences. Specifically, the success of a process ontology within physics is seen as a reason that process ontology should be taken as a viable scientific metaphysics. This is a powerful naturalistic reason to accept process philosophy, but leads process ontologists back to a monist metaphysics: *panta rhei*. The countercurrent that process ontologists face is clear. The influence of Cartesian mechanical philosophy has been an overwhelming tendency to see every natural event as "mechanical", organisms in particular being see as "machines", and so it is often thought that we should study machines to understand organisms. If process ontology is going to secure its status as the ontology of biology, it seems as though it will somehow have to displace the mechanical, "machine" conception. That is reasonable, but problems arise once this *panta rhei* conflicts with the machine based understanding of organisms.

Organisms, it is often said, are processes if anything is. Organisms are paradigmatic or obviously processes. The familiar or paradigmatic machines are not paradigmatic or obvious processes. In supporting a process ontology of organisms it makes some sense to distinguish them from machines, non-paradigmatic processes, but supporting process ontology as a metaphysics requires the view that machines are processes. Machines are processes, it must maintain on pain of inconsistency, since everything is. This is enough to show that the MOD cannot be established on the basis of the mere applicability of process ontology. This means that the process ontologist who wished to propose an MOD must shift to a method of arguing against machine ontology that is based on something else, beyond merely being or not being a process. What remains is for the process ontologist to argue that *machines are not the same sort of process as organisms*, and to offer specific MODs to this effect. Many different MODs can then be made on the basis of differences between sorts of process—Nicholson (2018) provides a volume of such differences.

But this strategy can also miss the mark, depending on which difference is proffered. One way that processual MODs go awry is being too general for their aim of advocating a process ontology. In an effort to displace a machine ontology of biology without giving up on *panta rhei*, process ontologists go through the trouble of offering an MOD interpreted within process ontology that could just as well have been offered without it. This is doubly bad for process ontology: if these MODs turn out false, they don't thereby offer any support to process ontology; if they turn out to be true, they are thereby just as supportive of other ontologies. The upshot of this is that "processual-MODs" can turn out to not be "specifically-processual-MODs".

For example, Nicholson (2018) locates an MOD in thermodynamics, arguing that "the second law allows us to see why [organisms] are fundamentally different from machines". He says the difference follows from the fact that organisms are systems that are "far from thermodynamic equilibrium" while machines are "equilibrium or near equilibrium" systems. Even assuming that this were a fundamental difference between them, this MOD could just as well be founded on (neo-)mechanism.[20] Indeed, non-equilibrium thermodynamics can be and often is presented as a theory about the activities of entities and as facts about (the rates of) chemical mechanisms. It might be right that thermodynamics is impossible to interpret within neo-mechanism—a very serious charge—but barring this, a thermodynamic-MOD sits outside process ontology in particular.

In searching for more specifically processual MODs it will help to look at those distinctions offered by advocates of processualism, but this in itself is insufficient. We are better

[20] I.e. one would need to assume that no organism were a near equilibrium system and no machine a far one, and assuming that non-equilibrium thermodynamics could itself be given a satisfying process ontological interpretation (there is no reason to assume it cannot).

off looking at distinctions offered specifically in the context of defences of those features of process philosophy thought integral to the contemporary (neo-)processual account of organisms. In § 4.1 I argued that these are (1) constitutivity, (2) hierarchical stabilization, (3) material flux, and (4) top-down stabilization. These will now be addressed in turn.

4.4.1 Constitutivity of Process and the MOD

The processes of an organism's life cycle are supposed to be constitutive of being that cycle—something cannot be the organism it is without also being a process. In connection with the constitutive aspect of processes for organisms (§ 1.2) we find two formulations: (1) organisms are not "properly understandable" in terms of one set of properties, but are ultimately processes, and (2) the life cycle (not its stages) is the organism. Without disputing this, we can see that there are also machines that cannot be understood in terms of a single set of properties, and for which a changing set of (sometimes cyclical) processes are constitutive of being that machine.

Life cycles (organisms) may appear stable during some period of observation, but this is a figment of brief observation. Dupré (2013) makes this dynamism of living processes explicit in a comparison to inanimate matter.

> At the cellular and molecular level it is even clearer that nothing stands still: a static cell is a dead cell... *[Biological processes] require some kind of internal change to continue to be the processes that they are*. A process, as mentioned above in the case of cells, needs activity to sustain it. One can imagine a rock undergoing no changes at all for, say, a minute without thereby ceasing to be a rock; a mouse in the same state of stasis is an ex-mouse.—Dupré (2013) p.30, my emphasis

Here Dupré highlights the importance of internal changes for the identity conditions for biological processes: being something or other requires changing internally. However, the same can be said of inanimate matter, under the right time and physical scale of examination. If a period of stasis significantly longer than a minute is considered, even a rock can lose its identity. Recall that the more general 'principle of process' (Whitehead 1929) sees the becoming of all entities as constitutive, suggesting that a rock in perpetual stasis ought also to cease to be the rock that it is. This is true for some: igneous rock cannot form in stasis.[21] Which sorts and durations of stasis are important is left somewhat open above. There are temporarily spore-forming microbes, desiccation tolerant animals, and plenty of

[21] *Formed by cooling of molten lava.*

aquatic animals that can be temporarily frozen solid without dying or losing their identity. Indeed, for some, going *without* periods of low-temperature or low-solvent stasis would be *uncharacteristic* of their lifestyles. And perhaps techniques may eventually exist for placing any multicellular organism in stasis artificially.

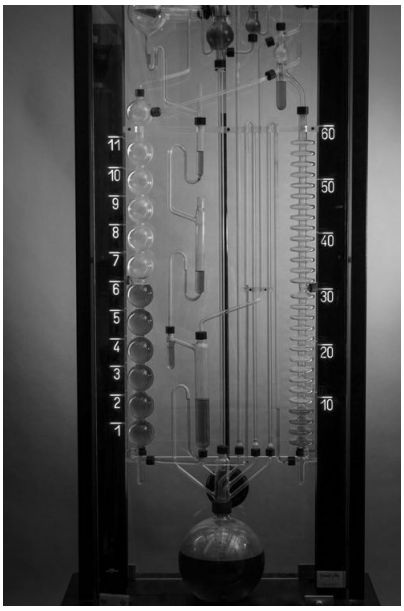


Fig. 4.1 *Time Flow Water Clock*, Artist: Bernard Gitton.

Of course, these really should not count against a processual understanding of organisms. Dupré need only point to periodic processes or larger timescales. The whole life cycle constituting a mouse cannot occur in stasis. A temporarily static cell may not be a dead cell, but the larger point about the constitutivity of life cycles remains. 'Static life-cycle' is a self-contradiction. Nonetheless, the identification of organisms by internal activities works against the MOD. If a mouse in stasis is an ex-mouse, then machines in stasis are ex-machines. If we want an MOD to follow from the identification of entities by their internal changes, then it must latch on to some present processes that make an entity active or functional, and not to their ability to come out of stasis unharmed. Both machines and organisms can enter and exit stasis. If a clock is a machine that more or less accurately tells the time, a clock in stasis is an ex-clock. The entire sets of effects that functionally constitute a machine, for cyclic machines, cannot occur in stasis; most require internal changes to continue to be what they are. A static machine is also an ex-machine. Machines can often be brought out of stasis, unfrozen with little damage, whereas organisms must either evolve systems for survising stasis or else be frozen and unfrozen under highly demanding conditions. However, some machines cannot be cycled through stasis without changing fundamentally—the thin glass tubes of Gitton's clock, pictured above, would shatter from the internal pressure of expanding ice, if frozen—just as some organisms are irremediably dead once static. The ability to move in and out of stasis is not an MOD.

For another example, a computer must be able to undergo a series of internal and precisely timed switching operations in order to be the machine it is—these processes are constitutive of computation. Moreover, that process philosophy can include aspects of computation within its wide descriptive and explanatory scope is, according to Seibt (2012), one of the powerful reasons to adopt it. We can look to computer "viruses" if a more clearly cyclical example is desired (see Ch.3 § 3.4.2). One would not understand what a particular virus was

without seeing how it copied itself and infected other computers—a copy of the "Love Bug worm" saved on an static hard drive is arguably not a virus at all. Processualists say much the same about biological viruses.

> Even viruses have life cycles... Viruses pass through an intricate sequences of stages as part of their life cycle. Some of these stages are highly stable (for instance, the virion stage, which is what most people have in mind when they think of viruses). Crucially, however, the very existence of these stable states can only be accounted for by referring to their role in the larger cyclical process that is the virus.—Dupré and Nicholson (2018) p.19

While perhaps most people have something comparatively more stable in mind when they think of computer viruses (such as the storage of the virus as data), if we wanted to explain why that particular data exists at all we would have to refer to an "intricate sequences of stages" that generate stable virus copies as part of a cycle of viral copying. Moreover, if we wanted to "properly understand" a computer virus, we could not do so without placing these stages in context of the computational processes that sustain their storage and numbers. Once we allow that processes can be constitutive of having a certain identity—such as being an organism, cell, or biological virus—we can likewise find examples of machines with precisely the same sort of constitutive relationship with internal changes.

Lastly, in keeping with Dupré's non-metaphysical inspiration for process ontology of biology, we can likewise draw from DST's relatively permissive notion of "developmental resource". Specifically, the notion of extended inheritance offered by DST provides a potential bridge between processual ontologies of biology and technology.

> DST views both development and evolution as processes of construction and reconstruction in which heterogeneous resources are contingently but more or less reliably reassembled for each life cycle [p.1]...[p.4] DST insists on a definition of inheritance that explicitly recognizes the wide range of resources that are "passed on" and are thus available to reconstruct the organism's life cycle. Some of these resources are familiar—chromosomes, nutrients, ambient temperature, childcare. Some are less familiar...—Oyama, Griffiths & Gray (2001) p.1-4

In a techno-society like our own, it is much more than "genes and childcare" that are transmitted between generations and serve to reconstruct developmental systems. Some of the developmental resources that are more or less reliably reassembled in life cycles are machines or artefacts (see Ch.7). We simply would not be the people we are without them;

an ox would not be the organism that it is without the plough. Likewise, the construction and design of mechanical systems themselves requires a range of resources far beyond merely schematics and material constituents. This should also point us to a "fundamental fact" about machines, that they cannot be understood in terms of one set of properties, such as the completed machine, but are rather constituted by processes.

4.4.2 Hierarchical Stabilization and the MOD

> An organism is not organized as a hierarchy of structures (as a machine is) but as a hierarchy of processes.—Dupré and Nicholson (2018) p.16

Here we suffer from a clash of primitive notions: 'process' is a fundamental category to process ontologists, thus undefined. Likewise, process ontologists are laboring against a structure-ism of sorts, so happen not to define 'structure'. This makes it initially difficult to see what kind of distinction is being offered. Presumably, the above could be reinterpreted as a processual claim: machines are a hierarchy of very stable processes, organisms not so.

Machines are indeed not paradigmatic processes because they are too stable, but it is explanations of stability instead of change that process ontologists offer. This would suggest that machines, as comparatively stable processes, require an explanation of how they remain stable "structures". Recall that it is precisely the apparent stability of certain organismal features that demands an explanation in terms of hierarchical stabilization.

> [T]he entities that form the hierarchy of biological ontology are not stable. They are, rather, stabilized over a very wide range of timescales, and the processes of stabilization are a fundamental part of the explanation of the activities of living systems.—Dupré (2013)

Since organisms display stability over some durations, presumably what is meant is that organisms are not intrinsically stable. This makes the distinction packed into the "structure vs. process" MOD clearer: organisms are stable because they are stabilized, machines just are stable processes already, without accompanying processes of hierarchical stabilization. Put another way, "[hierarchical] processes of stabilization" are not "a fundamental part of the explanation of the activities" of mechanical systems. In § 4 I will discuss hierarchical stabilization explanations within a general process philosophy of technology, but here concentrate on their role in the MOD.

The portion of this claim that is about organisms is reasonable. Much of the work explaining traits of organisms involves showing how these traits persist in the face of changing environmental conditions or internal degradation of components (§ 3.3, homeostasis being

a stabilizing process taking place at a lower hierarchical level), and against the constant introduction of deleterious mutations (§ 3.4, so-called "stabilizing selection" being a process at a higher level). When it comes to machines we can afford to be more particular.

Clearly some machines require little explanation of their stability but others are, like organisms, subject to effects which stabilize and destabilize their organization. To these effects, a machine may be more or less resistant, and more or less "actively" counteract them. Indeed, to hold to the idea that machines are somehow intrinsically stable, or stable simpliciter, goes against the central tenant of process ontology: that all processes are dynamic, "stable" only when stabilized. On whether these stabilizing effects are a "fundamental part" of explaining mechanical systems, this surely depends on how important we believe it is to have an explanation of why certain machines or mechanical forms remain stable, and which sorts of questions we ask about them (see § 4.4). In Ch.3, plenty of examples of self-maintenance in machines were offered, and surely some of those processes would be required to explain the activities of the maintained machine.

Perhaps the greater difficulty is interpreting what it means to be "hierarchical" in the case of machines. The theory of the biological hierarchy, though with some difficulties aligning organizational-levels with levels-of-selection, is largely uncontentious. Reproducing entities are nested, and successively larger inclusive biological objects have common names: gene, genome, cell, tissue, organism, species, etc. To my knowledge, no strictly analogous theory of the technological organizational hierarchy exists (see § 4.4). There clearly is a mereological hierarchy applicable to machines (Crilly 2013; 2015; see Ch.7). Machines consist of parts—stable processes—and collections of machines exist for longer or shorted times depending on intra- and interactions. Moreover, causal explanations of technology do involve different levels of explanation (Rapp 1983; see Crilly 2015 for an account of levels of functional explanation). In the following sections I will discuss processes of stabilization that are, in this sense, below and above the level of individual machines.

4.4.3 Material Flux and the MOD: Two Dogs and Empiricism

Dupré and Nicholson (2018) offer a thought experiment about hierarchical stabilization. It involves two dogs, one "living", the other robotic, and an extraterrestrial humanoid "whose mode of visual recognition was based on the enumeration of the material components that make up particular tokens of general types rather than on the identification of general types that are instantiated by particular tokens" [ibid p.16]. It suffices to say that this humanoid is somehow only attentive to material composition. The humanoid comes to earth, observes the dogs, presumably enumerates their material components at the molecular level, and returns again in a few years.

> Despite being in the presence of the same two dogs, the alien's cognitive apparatus is such that he is only able to identify the robotic dog and not the living one. From the alien's perspective, the living dog of the first trip has faded out of existence, and *an entirely different living dog* has taken its place. . . If one focuses on matter rather than on form and allows for a sufficiently extended period of time, the stream-like nature of macroscopic organisms becomes perfectly evident.—[ibid p.16]

This thought experiment would be a way of showing that identification by material composition suffers temporal limitations were we to assume that organisms (for instance) are subject to this kind of total material flux over a sufficient period. Instead, this is something we are supposed to accept as justified on empirical grounds, due to recent results about the turnover rates of tissues, cells and macromolecules. This justification is overstated: "In general, none of the parts of the organism are as old as the organism itself" (Dupré and Nicholson 2018, p.17), and "[the machine] serves as a channel that facilitates the exchange of material as fuel is converted into waste. An organism, in contrast, changes wholly and continuously" (Dupré and Nicholson 2018, p.146). This is incorrect, and overly cynical about the possibility of re-identification by material composition.

The biological details speak against this extent of material flux. Many cells are not replaced and certain macromolecular complexes therein persist as long as the organism does. The title of Wade's article, *Your Body Is Younger Than You Think*, is a double entendre. It also reports "a fact that explains why people behave their birth age, not the physical age of their cells: a few of the body's cell types endure from birth to death without renewal, and this special minority includes some or all of the cells of the cerebral cortex" [ibid]. Some of the parts of you that think are older than some of the rest of the body. Returning to the thought experiment, we would actually expect the extraterrestrial to see a more or less clumpy, vague, living dogly blur. A living dog would be at least in part re-identifiable until its death, no matter the elapsed time.

On the other hand, that a robotic dog would be totally re-identifiable surely depends on which robotic dog is under consideration. The "BigDog" robot released by Boston Dynamics is powered by gasoline, and so even a minimal runtime between extraterrestrial visits would have purged at least some of the material fuel constituents. Most robot dogs have replaceable batteries, or at least require the turnover of electrons (as all respiring dogs would). A more interesting example comes from the culture surrounding repair of the robot dog Aibo, released by Sony between 1999-2006.

In an article for the *New York Times* series *Robotica*, Soble (2015) follows Aibo owners and repair-people who replace the damaged parts of the (now discontinued) Aibos. Recently,

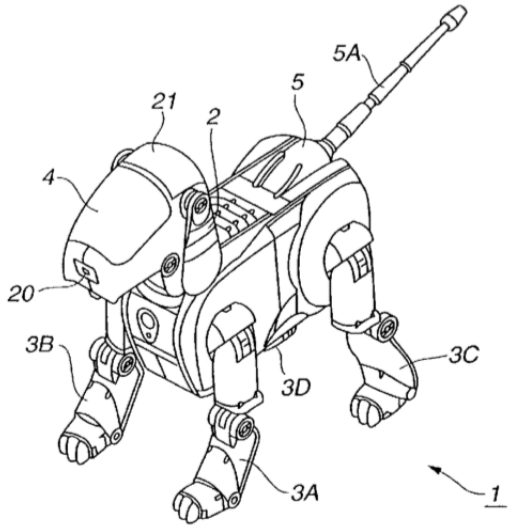


Fig. 4.2 Aibo, robotic dog released by Sony. Image from patent # US 7,200,249 B2

Sony decided to stop repairing Aibo, so owners have switched to private means for keeping their robotic companions in working order. This has been framed as a concern for the coming mortality of Aibo.

> The Aibo fell victim to company restructuring, as Sony sought to refocus on more profitable businesses. Still, Sony continued to repair Aibos until March of last year. But by then spare parts were becoming too scarce, the company said, forcing it to end the service and turn owners away. —Soble (2015), *A Robot Dog's Mortality*

There are Aibo today with a (partly) different material composition from what they had in 2006 and moreover, eventually, may differ completely in the material composition of their parts—unlike a living dog, which will likely retain parts of its cerebellum regardless of age. This may cause problems for our account of identity of Aibo over time, but no more so than it would for a living dog.[22] It is unsurprising that we should encounter a robot dog of Theseus problem on the basis of replacement of robotic parts. Repairing is a process and Theseus had a ship to fix, not a whale. The force of that through experiment is intended to encourage us to focus on "matter rather than on form" over a "sufficiently extended period of time". Though it is far clearer that a machine can change wholly in matter without changing in form, at least, in antiquity, clearer than it was that an organism could do so. Indeed, it is overstatement of

[22]Note that we are to assume a criteria of identity that is sufficient to say that we are indeed "in the presence of the same two dogs" (Dupré and Nicholson 2018) regardless of this flux

the differences between living and non-living things that makes the views of old-fashioned biological mechanists seem comparatively sensible, as Needham (1928) said,

> The mechanist, after all, never asserted as [is] against common sense that there was no difference between a stone dog and a real live dog ; he only insisted that the processes going on in the living dog were extremely complicated special cases of the processes known to occur in the inorganic world. —Needham (1928) p.34

This capacity to replace parts (see Ch.3 § 3.4.2) is thought relevant to distinguishing process ontology from both machines and neo-mechanism. Dupré (2012) is concerned with establishing a difference with new-mechanism and its emphasis on termination conditions (Ch.5). Organisms, Dupré argues, are processes that need not end in principle.

> Paradigmatic machines—cars, dishwashers, computers—consist of a number of parts, typically more or less rigidly connected. The constituent parts gradually wear out and the machine lasts as long as they are replaced piecemeal... One thing we add when we move to biological systems is that these, organisms for instance, constantly rebuild and replace their parts. Contrary to some versions of mechanism, there is no a priori reason why the process should end, hence no terminal condition. Lineages of organisms have no mechanistically inbuilt tendency to terminate.—Dupré (2012)

It is not clear that organisms can perpetuate themselves indefinitely by repair and replacement of parts—there is ongoing debate about the status of "programmed aging", i.e. an "inbuilt tendency to terminate" in some organisms (Goldsmith 2012).[23] Moreover, it is certainly the case that organisms "constantly rebuild and replace" some of their parts, but there are non-replaced and irreplaceable parts, and there is irreparable damage. An organism likewise must contend with gradual wear and lasts as long as it can repair and replace its essential parts, or they remain undamaged. Even if an ideal organism needn't terminate in principle, such beings are not the common stock. Likewise, some machines—such as Dextre, the space robot (Ch.3)—can replace and repair some of their parts, even though these may not be as common.

We need not conclude that organisms are in part a hierarchy of structures. It is hard to see how persistence of cerebral cortex cells over the course of a life-cycle implies that the

[23]The target seems to switch in the last sentence above: presumably the reason that lineages do not necessarily terminate is that they exist so long as any one of their members does, and that only requires the parts of the lineage to reproduce at least once before death. But lineages are plainly not organisms. If the claim where that life-cycles have no mechanistically inbuilt tendency to terminate, that would be an empirical hypothesis, and testable.

cerebral cortex is a "structure" as opposed to a process. We should, however, recognize that the "hierarchy of structures vs. hierarchy of processes" MOD, that is supposed to be established by turnover, fails. Some machines exhibit turnover processes at lower levels of organization and some organisms fail to be able to turnover their parts. Both are mortal; both face lower level flux.[24]

To address top-down stabilization, the last of the four aspects of processualist view of organisms under consideration here, it is necessary to have some conception of stabilizing selection in technology, and this will be taken care of in Ch.6. For now, I conclude by addressing hierarchical explanations within general process philosophy, since this is required for hierarchical explanations of stabilization of technology.

4.5 For A Processual Account of Machines

In the above (§ 4.3) I argued that many of the features thought central to a process ontology of biology are also enjoyed by some machines. Though characteristically organismal features of processes are perhaps "unparadigmatic" in machines, they are present, and this paradigm may shift. In this final section I will examine some features of process philosophy generally in an attempt to lay a foundation for a process ontology of technology. I will conclude with a diagnosis of the provision of processual MODs, itself in the spirit of processualism.

Notice that a mechanistic or machine based understanding does not, in itself, need to be placed at such a distance from process ontology. Indeed, most often, they are not as rigidly opposed as their advocates might like. Neo-mechanists (e.g. Machamer et al. 2000) and working biologists[25] are generally unafraid of admixtures of processual and mechanistic concepts. An extreme case can be found in Friedrich Rapp's (1981) definition of a "mechanistic viewpoint".

> (1) The root metaphor is no longer naturally occurring processes, but rather *mechanical processes*, artificially generated with the aid of appropriate apparatuses and instruments
>
> (2) The conceptual inventory for the description and analysis of all natural phenomena is no longer obtained from the 'higher' and complex organic processes, but rather from the 'lower' and simpler *inorganic processes*.

[24] Moreover, if this form of self-maintenance were really the essential bit distinguishing organisms from machines, it is hard to see why process ontology specifically hangs in the balance: autopoiesis or autonomy might have done just as well (Ch.3).

[25] Recall Kochanowski et al. 2013, § 2.4 above.

> (3) The *synthetic and teleological perspective*, which focuses on the final result of processes, *is now replaced by the analytic* investigation of the functional relationships between spatially and temporally contiguous states of affairs.
> —Rapp (1981) emphasis mine

Rapp (1981) defines mechanism essentially as a rejection of organicism and teleology (Ch.2) and final-causation, together with an emphasis on reductive (or, more charitably, bottom-up) explanations, analysis and artificial experiment. It would be hard to find a neo-mechanist today who denied any of these elements in their philosophy, though there are some who would wish to add something (e.g. holism, Ch.5; Austin 2016). Rapp's *Analytical Philosophy of Technology* could be read as a foundation for a process ontology of technology.[26] We can also take formulations like the above to indicate just how much of mechanism one can retain without having need to *oppose* this to process ontology. Dupré himself is also not "hostile to neo-mechanism", so long as its explanations are constrained by a general process philosophy view of stabilization. Neo-mechanist accounts can be successful, it is just that "the parts appealed to in machine explanation must be sufficiently stable over the time scale of the process you want to explain" (Dupré 2017). If Rapp had been writing a little later, one can easily imagine this constraint as a 4th feature of the mechanistic viewpoint.

Constraining scientific explanations via process philosophy leads to a pair of general principles for particular sciences. Let us call whatever non-process ontology deployed by a science or theory its "nascent ontology". Then, the move to a process ontology should obey at least the following constraints:

GP1 The entities referred to in the nascent theory are interpreted as processes.

GP2 The modes of causal explanation in the nascent theory should be augmented to provide explanations of stabilization of entities.

GP1 is the bare minimum for a process ontology. It involves attempting to construct a theory, or reinterpret a nascent one, such that static entities are analyzed as more or less stable processes. GP2 provides a further constraint on the explanatory structure of theories. A general process philosophy satisfying only GP1 might see some processes as stable simpliciter, but this would be to give up on an explanation of their stability. Instead, a good scientific theory should explain how stable processes become so, it should offer stabilization explanations—in effect, partially justifying entity-talk in nascent, pre-process,

[26] Though it would be a potentially equivocal undertaking to attribute too much genuine process philosophy to Rapp on the basis of his preference for the term 'process'.

theories. Dupré provides such a justification when explaining, in processual terms, the conditions under which mechanistic explanations can be successful.

> Mechanistic explanations will be successful only to the extent that the constituents identified are sufficiently stable on the timescale of the phenomenon under investigation.—Dupré (2017) p.3

The move to hierarchical stabilization explanations among process philosophers can also be justified by appeal to GP2. Many modes of explanation in biology deploy some concept of a "biological hierarchy", so a mature process philosophy of biology should be able to augment these nascent explanations to employ the biological hierarchy in explanations of stability. Much of the work of process ontologists of biology is directed to show that many hierarchical biological explanatory strategies are processual already (e.g. homeostasis, stabilizing selection) and thus that process ontology is not imposed on biology, but instead arises naturally from a "proper understanding" of it. Likewise, whatever hierarchical modes of explanation exist in our nascent theory of technology should, by GP2, be augmented to provide stabilizing explanations.

The biological hierarchy can be interpreted as a form of hierarchical classification, where "levels" like macromolecule, cell, species and family serve book-keeping roles within biological theory. It can also be interpreted as a hierarchy of levels of selection, where levels from macromolecules to higher-taxa (e.g. genus, kingdom) are interpreted as units of selection.[27] Finally we can consider a mereological or organizational hierarchy, where inter-level relationships are part-whole or part-individual. Constructing a processual analysis of these hierarchical modes of explanation is well underway (many details of which are covered in sections above and in Dupré and Nicholson 2018), but incomplete. That is simply because process ontology cannot have a complete reinterpretation of an unfinished theory.[28] Lacking a processual account of these modes of explanation, process ontology of biology is only partway to satisfying GP2 for hierarchical explanations.

The problem for a process ontology of technology is, then, identifying the existing analogous modes of causal explanation in technological theory and augmenting them to supply stabilization explanations. Even holding to the standard set by work in process ontology of biology, we should be lenient enough to allow these explanations to only partially

[27] Typically the only macromolecules under consideration as levels of selection are specific fragments of the genome (genes, cistrons), but there is increasing recognition that specific proteins (prions) and even RNA fragments can be selectable units.

[28] Post facto, it is interesting to look at Dupré's (1995) *The Disorder Of Things* as an attempt to spell out the consequences of a process ontology for the classificatory hierarchy. Dupré now sees many of the substantial points about classification in Dupré (1995) as explained by a general process philosophy (Dupré, personal communication 2019, also see Dupré 2018).

satisfy GP2. For the specific case of the MOD, that requires identifying the role of top-down hierarchical explanations. For instance, explaining the success of some machines will require identifying the causally relevant features of their design process, which will surely require an explanation for why the parts of a machine remain sufficiently stable over the course of their functioning. Such a stabilizing explanation might come "from below", via a materials science explanation of the durability of design components or via in-built self-repair mechanisms. But it can also come "from above", via explanations of other-maintenance, repair processes or the stabilizing effects of collections of machines (e.g., the effect of anti-virus software on the relative stability of a given computer in a network).

That the hierarchical stabilization explanation for technology involve selection requires digression to a discussion of whether certain technological processes are analogous to natural processes of evolution by selection. To my mind, here is nothing wrong with giving a "bare-bones" selectionist explanation of technology change; disagreement only seems to arise as to the significance of this explanation. That is enough to imply that there ought to be "stabilizing selection" during technology change, however (in)significant in extent or explanatory scope. That is, insofar as selectionist explanations apply to technology at all, the fact that potentially deleterious or damaging changes are introduced into technology "from below" implies that stabilizing selection will play some role. An account of stabilizing selection in machines will need to come from an account of technological evolution. I save a sustained treatment of this contentious topic for the penultimate chapter (Ch.6).

4.6 Conclusion: Poorly Defined Targets of Process Philosophy

Contra Dupré and Nicholson (2018), GP1 demands that we see machines not as a hierarchy of structures but as a "hierarchy of processes". One way of seeing GP2 is as a requirement that we use this hierarchical understanding to explain why it is so tempting to see machines as "structures". Recognizing this takes us towards both an understanding of, and possible reconciliation with, attempts to provide an MOD on the basis of process ontology. I conclude with a process philosophical diagnosis of the provision of processual MODs.

Austin (2016) offers an hypothesis as to why process ontologists have directed arguments against mechanism.

> [T]he contemporary dialectic against organisms qua ontologically composed of mechanisms appears to be often focused on objections to organisms qua machines.—Austin (2016) p.659

He goes on to say that this is reasonable, justified by the claim that machines are "ontologically outmoded" and "outmatched by our contemporary science" [ibid]. This sociological hypothesis about the origin of anti-neo-mechanistic discourse might be correct in some cases—it surely is for Woese (2004) and Nicholson (2018)—but Austin's justification is not. Machines have been around in philosophical discourse much longer than neo-mechanism and their development is a very large part of our contemporary science. They are not outmoded, since they are consistently outmoding their previous incarnations. Instead, the best processual critiques of neo-mechanism are just the same as those for machines: the (apparent) relative stability of the "entities" involved in explanations.

Since machines are stabilized, we can explain their activities without reference to their processual nature at some time- and organizational-scales. This is also true of organisms, but our interest in explaining the activities of organisms has delved into phenomena at different scales—physically smaller metabolic processes and temporally extended evolutionary ones. So progressively more processual understandings of organisms became important in biology. Since it is sometimes acceptable to explain mechanical activity without an explicitly processual understanding (since machine parts are more often stable over the duration we wish to explain), and since we are more inclined to provide processual understanding of organisms, we are thus disinclined to provide a mechanical explanation of organisms. I suggest that this is the true reason why process ontologists of biology tend to construct MODs as if they supported general process ontology, when in fact they do not.

This could have gone differently. Instead of arguing, given process metaphysics as the right view of organisms, that the analogy with machines must be rejected, one could have argued that since machines are much like organisms and a process ontology is the right view of biology, that there is reason to expect it would be the right view of technology also. That is, we might just as well have sought out a process ontology of technology to parallel the process ontology of biology. This would not contend with the old machine metaphor but update it, in light of a naturalistic ontology of organisms and examples of comparatively more processual contemporary machines.

In similar fashion, the following chapter (Ch.5) shows that neo-mechanism has also become unnecessarily embroiled in the MOD, by both its critics and advocates.

Chapter 5

Mechanism and the Machine Organism Distinction

Abstract

In this chapter I examine MODs advanced in the context of neo-mechanism, both by advocates and critics. Like process ontology (Ch.4), neo-mechanism offers an ontological framework for analyzing both machines and organisms, so some further MOD must be provided within that framework. I examine two cases, one by critics of new-mechanism and one by an advocate. Skillings (2015) and Godfrey-Smith (2016) offer similar critical views: that neo-mechanism's reliance on the non-stochastic nature of machines makes mechanism inappropriate as an analysis of organisms. I argue against this firstly by showing that new-mechanism is not strongly committed to non-stochastic mechanisms, and secondly by showing that there are many machines that operate in a fundamentally stochastic way. Austin's (2016) MOD is that neo-mechanism can account for the holistic phenomena observed in organisms, while the earlier, "machine-based", mechanistic ontology cannot, since machines are not holistic. I show that Austin is mistaken, since even on a fairly restrictive account of what it takes to be a "holistic phenomenon" there are holistic machines. This chapter concludes my metaphysical engagement with the MOD.

Introduction

This chapter, like the previous, addresses the MOD as a metaphysical issue. A variety of metaphysics, Mechanism, bears on the relationship between machines and organisms. Mechanism is a diverse collection of views. Running from the "Mechanism" of Descartes (i.e.,

mechanical philosophy) to the "New-Mechanism" of the new mechanists (e.g., Machamer et al. 2000; Bechtel 2011), there is perhaps not a single unifying thread. I will bring some clarity to this group of views by setting apart the different senses of 'Mechanism', specifically as they pertain to the MOD. I will write 'Mechanism' when the broad, diverse, amorphous collection of views is intended and will define subsequent refinements of this term as necessary.

To critics, appealing to the "mechanical" or "mechanistic" nature of organisms seems to incorrectly or unjustifiably equate or reduce them to machines (Moore 2012; Skillings 2015; Godfrey-Smith 2016; Austin 2016). That this equation or reduction is problematic presumes some kind of background MOD. Thus, many criticisms of "Mechanism" tend to advance or rely upon MODs. This reliance is particularly apparent in criticisms of "new-mechanism", a view stemming from Machamer et al. (2000). For its proponents, new-mechanism does not confound organisms with machines. Rather, new-mechanism offers a metaphysical framework for analysis of both machines and organisms. New-mechanism thereby almost completely distances itself from machine-based analogies and metaphors. But critics of new-mechanism see it differently: they think that a connection to some of the properties of machines fundamentally limits new-mechanism, and prevents it from being equal to the task of analyzing organisms. I first consider how contemporary critics and advocates of new-mechanism rely on MODs (§ 5.1).

Here critics of new-mechanism go astray. By confining their understanding of mechanisms to those that are typically realized in a narrow class of machines (clocks, clock-work), they fail to understand the true expressiveness of the new-mechanistic framework and thus attack a straw-man. Specifically, Skillings (2015) and Godfrey-Smith (2016) criticize new-mechanism on account of its supposed inability to account for some of the stochastic micro-phenomena of organisms—taking the ribosome as an example. They attribute this to the relative importance of "regularity" and "productive continuity" in new-mechanisms, using disanalogies between machines and organisms in support—Skillings taking clocks as an example, while Godfrey-Smith prefers Leibniz's mill. Whether this disanalogy is rhetorical or doing logical work is somewhat obscure. I begin by showing that, if it is a rhetorical device, then it is dispensable and false on independent grounds. Next I show that, if the disanalogy with machines is doing logical work—if it is indispensable—then it is false on account of a failure to consider the true diversity of mechanisms, and the machines and organisms that they apply to.

Remarkably, both new-mechanists and their critics often point to the very same MOD: that organisms are "holistic" and machines not so. In previous debates between Mechanists and Organicists, this holistic nature of organisms was a common Organicist retort to the emphasis on reducibility on the part of Mechanists. Today, this criticism has been taken on

by Austin (2016) as speaking, still against Mechanism, but also for new-mechanism. Austin (2016) holds the view that new-mechanism can accommodate "holistic" phenomena, which cannot be done with a machine-based Mechanism. Whatever one's view on holism, this is a specifically new-mechanist MOD and fits well with the overall new-mechanist project of offering a framework for analyzing organisms. Nonetheless, this unnecessarily associates justifications for new-mechanism with an MOD based on holism. In keeping with previous chapters I argue that, even on a fairly restrictive account of what it means to be an holistic phenomenon in organisms, there are also holistic machines (§ 5.2).

That is the extent to which I will discuss contemporary involvement of Mechanism in the MOD. Nonetheless, there is a temptation that should be forestalled afterwards. It is tempting to conclude that, while modern metaphysical theories have been reformulated sufficiently to escape criticism by MODs, that their predecessors at least must have been susceptible to these machine disanalogies. Put another way, although perhaps new-mechanism is safe from disanalogies with (modern) machines, the Mechanism of the early 20th century, or the mechanical philosophy of Descartes at least, must not have been. I conclude this chapter, and my treatment of metaphysical MODs generally, by arguing that in fact Mechanism has *never* been about reduction to machines (§ 5.3). Claims that organisms are "mechanical" within Mechanism turn out to be contingent hypotheses that neither equate organisms with machines nor establish an MOD.

5.1 MODs *Against* New-Mechanism: Are Machines even Mechanical Anymore?

The "new-mechanists", stemming from Machamer et al. (2000) offer a novel metaphysical framework for thinking about the ontology of organisms. New-mechanism provides an ontology that has received substantial attention from the philosophical and scientific communities alike. To some, it seems to come with implications about their relationship with machines. This section outlines new-mechanism and some of the criticisms of it that rely on MODs.

New-mechanism entails at least an ontological commitment to both entities and activities, with the addendum that mechanisms are arrangements of the former that are connected by the latter that together show how some phenomena to be explained proceeds from some initial to terminating conditions (Glennan 1996; Machamer et al. 2000; Austin 2016; Craver and Darden 2013). The philosophy of new-mechanism is essentially an analytic project concerned with providing an analysis of familiar mechanistic diagrams that is consistent with a broadly materialist metaphysics. There are different definitions of mechanism coming

from different contributors to the field, but it seems natural to group them together not only because they label themselves similarly, but because their concepts are roughly the same. Consider parallel definitions from Glennan (1996), Machamer et al. (2000) and Bechtel (2011),

> A mechanism underlying a behaviour is a complex system which produces that behaviour by the interaction of a number of parts according to direct causal laws.
> —Glennan (1996), p.52

> A [basic] mechanism is construed as generating a phenomenon (e.g., protein synthesis) through a start-to-finish sequence of qualitatively characterized operations performed by component parts.
> —Bechtel (2011) p.534

> Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions.
> —Machamer et al. (2000) p.3

New-mechanists have good reasons to see mechanisms as definable independently from machines. 'Mechanism' for new-mechanists is not a metaphysicalizing of the notion of machine. It may have its intellectual antecedents in thinking about machines, but today it is a fledged metaphysics of its own, unfortunately wedded to 'machines' etymologically. It is perhaps better to think of it as an entities and activities-ism, for it is those rudiments, not familiar machines, that serve as its fundamental notions. They have overcome one of the significant obstacles of earlier Mechanisms: being confused for a naïve view wherein mechanisms are simply *equated* with machines.

New-mechanism nonetheless has critics who believe that its close association with machines makes it a poor or insufficient theory of organisms. Godfrey-Smith (2016) and Skillings (2015), both relying on the discussion of Moore (2012), argue that many of the features of new-mechanisms are too restrictive to apply to organisms and their parts. Both draw special attention to emphasis on *regularity* in the definitions that new-mechanists supply for mechanisms, arguing that this limits these accounts—without modification—from accounting for the *stochastic* nature of organisms. I will examine this stochastic MOD first, as it leads naturally to a renewed interest in holism.

Citing Moore's (2012) discussion of the ribosome as a molecular "device", both Skillings and Godfrey-Smith point to the comparatively random, probabilistic, or stochastic nature of its movements as opposed to the "productive regularity" of new-mechanisms. Both offer their

critique of new-mechanism alongside/using disanalogies with macroscopic machines—clocks and Leibniz's mill[1]. Skillings' discussion is the more compact, presented as follows,

> Think of a machine like a mechanical clock. The workings of the mechanical clock are explained by looking at how the physical parts of its mechanism interact... [here follows a description of the movements of a clock]... The movement and interactions of the parts of the watch explain how the watch works... The parts of a protein, like a ribosome, do not stand in the same relations as the parts of a mechanical clock. The mainspring of a watch can reliably turn the gear train. Macromolecules immersed in liquids are at the mercy of frictional and thermodynamic forces... the ribosome is randomly pushed into different conformational states by external forces, rather than moving in sequence according to the activity and interactions of its parts. Thus, all the functionally significant activities of the ribosome, both internal and external, in protein synthesis are probabilistic. —Skillings (2015) p.1145

The intended take-away is clear: "The mechanism of protein synthesis does not fit neatly into the basic [new-] mechanistic account. There is no productive continuity between stages, where earlier stages produce later stages. Rather, the mechanism proceeds stochastically" [ibid, pg.1146]. Earlier, Skillings points to this as one (of three) criticisms of the adequacy of new-mechanism for explaining biological processes. New-mechanisms apparently come with such a strong requirement for "regularity" that they are unable to explain stochastic biological processes, such as some of those involved in protein synthesis.

First, for my purposes it is helpful to separate the critique of new-mechanism as much as possible from the claims about machines. Above, when Skillings says for us to "Think of a machine like a mechanical clock", presumably what is at issue is the likeness of clocks alone to new-mechanisms, so that the logic of his criticism can be put something as follows: (1) to be explanatory, new-mechanisms must have regularity and productive continuity (as for example clocks do), (2) some biological processes do not have these features but instead have the contrasting or opposite features of stochasticity and/or productive discontinuity, therefore (3) some biological processes cannot be correctly explained by new-mechanisms. On this reading the (dis)analogy with machines broadly, and with clocks in particular, is idle.

That said, (1) is dubious. Skillings (and Godfrey-Smith) have taken the "productive continuity" of Machamer et al. (2000)—which is an oddly defined term of art—and interpreted it in narrow physical terms as non-stochastic, non-probabilistic, "deterministic" etc., when the

[1] I will primarily follow the discussion in Skillings (2015) since Godfrey-Smith (2016) is concerned with the bearing of this issue on the emergence of mental properties. Though I have dealt with this latter issue separately (Brunet and Halina *2021*).

interpretation actually offered by new mechanists relates to the completeness of mechanistic descriptions and the existence of explanatory gaps. Machamer et al. (2000) say,

> Complete descriptions of mechanisms exhibit productive continuity without gaps from the set up to termination conditions. Productive continuities are what make the connections between stages intelligible. If a mechanism is represented schematically by $A \rightarrow B \rightarrow C$, then the continuity lies in the arrows and their explication is in terms of the activities that the arrows represent. A missing arrow, namely, the inability to specify an activity, leaves an explanatory gap in the productive continuity of the mechanism. [ibid, p.3]

Likewise "regularity" is a Humean property and not equivalent to determinism or non-stochasticity.[2] It is closely connected to "productive continuity" (indeed, nearly synonymous with it) as follows: "The regularity is exhibited in the typical way that the mechanism runs from beginning to end; what makes it regular is the productive continuity between stages" [ibid]. What the new mechanist has issue with is not stochastic activities, but absent activities. If a mechanism for protein synthesis were missing an arrow, for instance the arrow representing the translocation step (wherein an aminoacylated tRNA moves from the A to P sites following hydrolysis of ATP), then it wouldn't be complete or, perhaps, it would be cut into two separate and complete mechanisms, one leading up to and another following translocation.

It is difficult to see why a new mechanist would have any trouble providing "stochastic mechanisms" by specifying "stochastic activities", and thus difficult to see why a new mechanist would have to provide mechanisms that are anything like those of clockwork. This is standard practice in the drawing of biochemical mechanisms: arrows are written with the thermodynamic (statistical) conditions under which they proceed at an appreciable rate. Since Machamer et al. (2000) and other avowed new mechanists are explicitly engaged in the project of providing an analysis of mechanisms from molecular biology and biochemistry, it would be very odd for them to be committed to a strong view of 'regularity' wherein activities were never stochastic. Consider the following mechanistic diagram (Fig.5.1) from McAdams and Arkin (1997), with the accompanying description.

> Each competition is an independent event with a probabilistic outcome. A transcript is initially in state 1 and thereafter in one of the five states shown in B.—[ibid]

[2] Although Machamer et al. (2000) disagree that intelligibility is reducible to regularities, offering the contrary view that it is explanations (of course, in new mechanistic terms) which "produce" or "sustain" regularities.

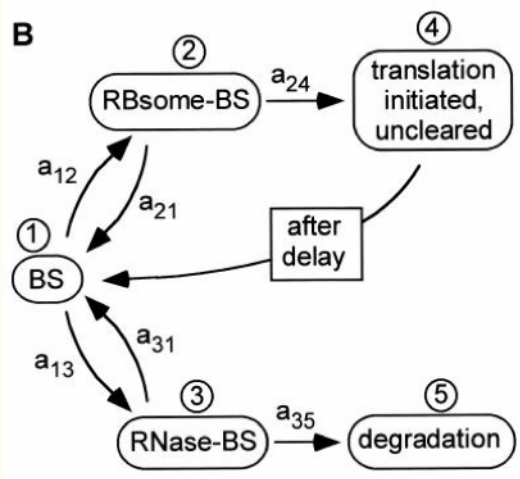


Fig. 5.1 Stochastic mechanism presented in McAdams and Arkin (1997), in section "Statistics of the Number of Protein Molecules Produced per Transcript". Here the activities $a_{i,j}$ indicate the probability of a transition from state-i to state-j for states 1-5.

The idea that new-mechanisms are wed to determinism insofar as (familiar) clocks are is simply untenable given the state of mechanistic diagrams within biochemistry. Moreover, these sorts of mechanistic depictions / interpretations are not uncommon or marginalizable: GoogleScholar returns 7, 520 results containing the exact phrase 'stochastic mechanism'.[3] Skillings's (2015) overall claim is nonetheless quite acceptable: that whether a biological process can be explained by new-mechanisms is more-or-less. Though, admitting the ubiquity of mechanisms with stochastic activities in scientific practice, the metaphysical picture offered by new-mechanism can probably account for more, rather than less.[4]

All of this sits aside the clock-analogy, making the place of this (dis)analogy within criticism of new-mechanism all the more questionable. Notice that there is an ambiguity in Skillings' analogy: if the reference to clocks is not meant to single out clocks in particular, but instead uses clocks as a representative of the entire class of machines—as is a viable interpretation of the sentence "Think of a machine like a clock."—then the analogy is making a far stronger claim about machines. If the reference to clocks is not idle, that is, if it forms an integral part of the argument, then the logic should be something as follows: (1') new-mechanism can correctly explain only those things which are like machines in their possession of productive continuity and regularity (and clocks are an arbitrary representative of this class), (2') some biological processes are not like machines in these respects, since

[3] January 2021

[4] Notwithstanding processual criticism discussed in Ch.4. My aim here is not to argue for new-mechanism.

they are not like clocks in these respects, therefore (3) some biological processes cannot be correctly explained by new-mechanisms.

Again, we may have reasons within new-mechanism to reject (1'), but even outside of a defence of new-mechanism we have reason to reject (2'): some organisms are just like machines with respect to their continuity and regularity, since some machines are unlike clocks. It is this latter point that concerns me most, since unless we interpret the analogy to machines as dispensable, idle, then we are pushed into an interpretation wherein an MOD, based on productive continuity and stochasticity, serves as a justification for critiques of new-mechanism—and my aim is only to defend new-mechanism insofar as criticisms thereof rely on or imply an MOD.

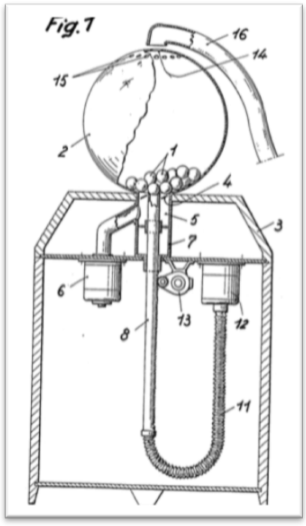


Fig. 5.2 Lottery machine for selecting numbers printed on balls (1) using fan (4) and enclosed vessel (2) to generate randomized selection. US patent # 4,961,578

Put another way, the starkness of the contrast between biological processes and new-mechanisms is an artifact of the selection of the artifact: clocks are not much like ribosomes, but that says little about the overall likeness of organisms or ribosomes to machines generally speaking. Indeed, a disanalogy with clocks says even less about what a new mechanistic account of protein synthesis should amount to. If instead we considered the comparison of ribosomes to machines that do include elements of randomization or stochastic processes, such as the lottery machine (Fig.5.2), the contrast becomes harder to establish. Merely being probabilistic or stochastic is not a distinction between machines and biological parts. If this disanalogy with machines is going to serve as criticism of new-mechanism, we will need to include more aspects of new-mechanisms in the disanalogy. But the Moore-type counterexamples actually become weaker as we do so. For instance, if we include the idea that ribosomes are "randomly pushed into different conformational states by external forces, rather than moving in sequence according to the activity and interactions of its parts" [ibid], then we can find watches with just such stochastic functions.

Admitting that being "randomly pushed into different conformational states" is itself a matter of degree, the ribosome is surely not completely deformed during normal functioning.

Evolution is remarkably good at generating molecular structures that tend to produce some effect under normal physiological conditions of excitation and external forces—thermodynamic and electric, but also Newtonian—by exploitation of so-called "molecular ratchets". In ATP-synthetase, the stochastic flow of H^+ ions through one component results in a regular circular motion of another, due to a similar molecular ratchet. More or less random conformational disturbances (or ion flux across a gradient) are transformed into productive and regular movements. This is remarkable, but not unique to biology—the name of this sort of process of course derives from the name of a common tool.

In the self-winding clock (Fig.5.3), part (5) is able to move randomly in either direction. It does so according to the movements of the wrist—when removed from one's pocket—on which it is fastened. As it does so, the parts (8a-b) tend to result in an overall counterclockwise movement of part (11), which itself turns (13) and contributes torque to a spring. The "regular" unwinding of the spring powers downstream movements of parts of the clock that require a constant output of energy. This is an effective way of transforming the stochastic movements of the entire clock into regular movement of the dials.[5] Moreover, if it is the nature of these external forces that is at issue, suffice it to say that the dependence of the watch on Newtonian gravitational forces is a contingent aspect of the design of this particular machine. One can just as well construct mechanical ratchets that operate on thermodynamic or electrostatic principles, to be used in environments with sufficient external forces of that variety. Tide-turbines, rotating wind-mills or fastening magnets to the poles of component (5) would all do just that. Again, contra (1'), it is hard to see why a new mechanist would have trouble devising a new-mechanism for this clock—the above image (Fig.5.3) is nearly all that is required.

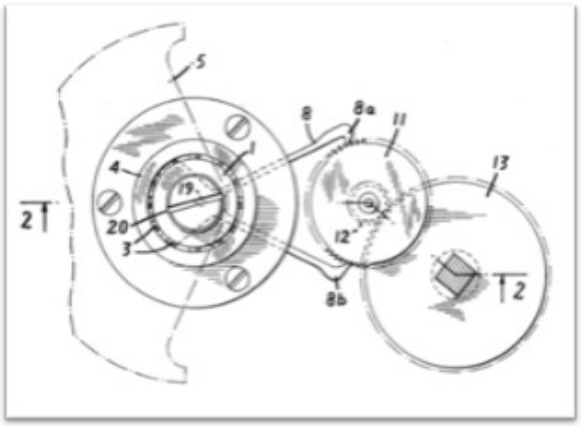


Fig. 5.3 Self-winding clock.

[5] It would be nice if we could here extend the analogy by saying "just as the structure of the ribosome is an effective way of transforming stochastic thermodynamic and electric forces into the movement of the bound mRNA" but this is actually achieved by the *perhaps even more new-mechanistically tractable* process of GTP hydrolysis driving protein conformational-state change. In short, the ribosome may be deformed during normal functioning, but this is far more accidental than the deformations during the functioning of molecular ratchets like ATP-synthetase, or the self-winding watch.

This is enough for my purposes: in arguing against new-mechanism on the basis of disanalogies between machines and the parts of organisms, critics either use these disanalogies in a dispensable way—in which case they should be dispensed with—or rely on MODs that turn out false on account of mechanical counter-examples. On the other hand, even if (1) and (1') turned out to be fatal to new-mechanism, i.e., these Moore-type cases are counterexamples to new-mechanism after all, then new-mechanism would also fail to be a theory of explanation in machines.[6] Interpreting productive continuity as non-stochasticity entails that new-mechanism cannot explain stochastic machines, that it cannot handle lottery machines, *inter alia*. If it cannot handle ratcheting of random conformational changes due to external forces then it cannot handle self-winding clocks—or sailboats. If it really were inapplicable to organisms due to their stochastic behaviour at the molecular level, it would also be inapplicable to machines, and so does not establish an MOD.

More is at issue. In Godfrey-Smith (2016) the biological predominance of randomness and trends ("ratcheting") is connected explicitly with the critique of mechanism via holistic phenomena.

> If we were observers of a living system at an intracellular scale, we would see some "parts that push one another," but not in the manner of macroscopic machines, and we would not only see pushes. We would see a storm of activity biased by charge and shape, generating partially random walks that, on average, tend in orderly directions. The processes are more causally holistic, noisier—more a matter of "herding molecular cats"—than a push-pull model allows.
> –Godfrey-Smith (2016) p. 189

The opening remarks of Machamer et al. (2000) contain the following cautionary remark: "one should not think of mechanisms as exclusively mechanical (push-pull) system" [ibid, p.2], and so we can safely say that this disanalogy with macroscopic machines misses the mark. New-mechanism is just not restricted to push-pull systems. Moreover, Godfrey-Smith does not offer an analysis of what is 'causally holistic' about biochemical processes. That is probably for the better: the analogy with "herding molecular cats" is derived from explanations of statistical mechanics, so it seems that "causally holistic, noisier" refers back to the statistical, stochastic nature of biological processes. We will require a more substantial reading of holism if it is to provide a special MOD. However, it is somewhat unsurprising that a contemporary *critic* of new-mechanism would take up a concept (holism) that was

[6]Moreover, new-mechanism would also fail even to be an account of textbook mechanism, such as that of the Krebs cycle. This is primarily because new-mechanisms are apparently confined to sequential processes (Skillings 2015), which the Krebs cycle is not. Although, one wonders why a cyclical process is not simply two (sequential) new-mechanisms jointed head-to-tail-and-tail-to-head.

historically reserved for organicists. In the following section I will show how an MOD based on holism has been taken up in a contemporary *defence* of new-mechanism, and argue that there are holistic machines.

5.2 MODs *For* New-Mechanism: Holism

5.2.1 Has New-Mechanism Become Retro-Mechanism?

Unlike the previous chapter, where it was advocates of process metaphysics that tended to supply or rely on MODs, up unto this point I have discussed cases where arguments *against* Mechanism have done so. I put this first because these criticisms pick up on a common theme: stochasticity is *prima facie* problematic to new-mechanism and seems to imply a need for a holistic treatment. In this section we return to the previous tendency: the use of an MOD when advocating *for* new-mechanism.

Consider the perspective of Austin (2016), an avowed new-mechanist, on the rise of process ontology.

> The contemporary dialectic against organisms qua ontologically composed of mechanisms, appears to often be focused on objections to organisms qua machines... A contemporary mechanistic ontology is not an ontology of machines, and the conflation of the two seems to function as an implicit cause of much of the process theorists' ire.—Austin (2016) p.16

Austin is right that new-mechanism is not an ontology of machines, but some of his reasons for this claim are distinctly organicist (more in § 3). On Austin's (2016) view, the dialectic of organisms qua machines implies the "rejection of holistic phenomena" [ibid], while that of organisms qua mechanisms does not. Insofar as an analysis of organisms requires holism, the new-mechanist can supposedly provide what machines lack.

Unlike Godfrey-Smith (2016), Austin (2016) provides two examples of organismal holistic phenomena: "self-replication and self-regulation" (p. 659). These have been addressed previously as features of autonomous capacities and intrinsic teleology (Ch.2-3), so I will not rehearse any specific objections to them here. Perhaps they are "holistic". If so, then Ch.3 shows that there are holistic machines. However, Austin's main project is showing that new-mechanism can accommodate another yet unaddressed ("higher-order") phenomenon: developmental modularity.

> [For] the higher-order phenomenon associated with developmental modules—namely, their multiple realizability... we require an ontology that can account

> for the persistence of those modules throughout substantial variation in the constitutive collection of entities and activities... entities and activities are no doubt of central importance to our contemporary conception of 'mechanism', but so is a particular dynamism.—Austin (2016) p.12

Austin argues that, since new-mechanisms can display the right kind of dynamism with respect to variation in constitution, they can allow for multiple realizability (of a module) over many different constitutions. Austin (2016) requires this to show the sufficiency of a mechanistic view with respect to the ordinary variation in constitutions of organic parts—that certain developmental events obtain regardless of variation in underlying genetic or biochemical entities. It is clear why a new mechanist would need this: if the framework for new-mechanisms cannot accommodate underlying variation during development, then so much the worse for the empirical adequacy of new-mechanism.

This is an important point to defend when advocating for a mechanistic view of organisms, but does not directly imply any particular MOD. So far Austin has only, to my mind rightly, argued that new-mechanism can account for variation in the constituents of multiply realizable biological modules. In this respect Austin's project is much like mine (§ 5.1), he aims to show that new-mechanism suffices for the analysis of a sort of phenomena that its critics say it does not. Although, he endorses new-mechanism. To further distinguish new-mechanism from the old, Austin claims that the machine conception cannot account for this: "unlike machines, contemporary mechanisms... allow for holistic phenomena arising from the collective activity of parts" (Austin 2016).

This is a drastic change in advocacy for mechanism. Organicist critiques have apparently been taken on board and incorporated. Holism, as an MOD, has been brought into the fold as a reason in favour of new-mechanism and machine analogies are explicitly the target of mechanistic criticism. Moreover, the reductionist hallmark of mechanism is explicitly excluded.

> 'Machine', being exhaustively dissectible into sets of entities whose activities can be studied in isolation, are strongly associated with the philosophical project of reductionism and the rejection of holistic phenomena... [U]nlike machines, the activity of contemporary mechanisms does not consist wholly in step-wise successions through linear series of clockwork-like connectives among their parts.—Austin (2016) p.16

Whether new-mechanisms can actually live up to these revisionary desiderata is beside the point of the present connection. What is interesting for me in Austin's (2016) approach is that holism is presented as an MOD and cast within a new mechanistic ontology. His

rejection of the "organism qua machine" stems from a rejection of "machine qua holistic mechanism". To show that this MOD also fails I turn to arguing that Austin's reasons for rejecting the "organism qua machine" furthermore stem from a limited view of machines and organisms: organisms are assumed to all be equally holistic and clocks are treated as representative of the entire class of machines.

Admittedly, the parts of most organisms do not much resemble clockwork. But there are organisms with parts that do. Consider the insect *Issus* that uses interlocking gear-like structures in its legs to coordinate movements (Burrows and Sutton 2013).[7] Here it is the diversity of organisms and their mechanisms that is being neglected: "We usually think of gears as something that we see in human designed machinery, but we've found that that is only because we didn't look hard enough" (Sutton, interview).[8] Evolution by natural selection rarely produces structures with such functional likeness to our simpler machines, but of course there is nothing naturally preventing it from doing so on occasion.

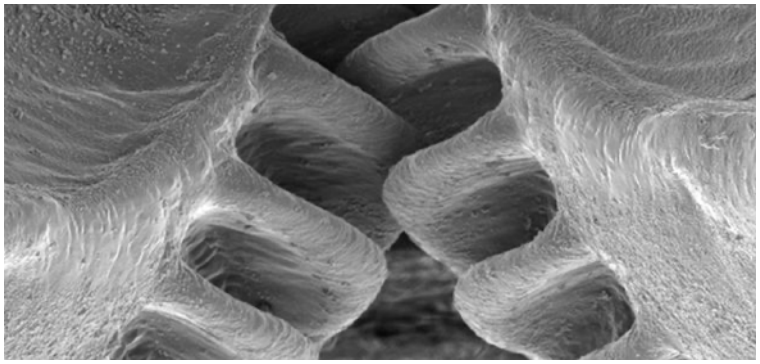


Fig. 5.4 Gear-like structure used to time movements in *Issus*.

This case is more than a quick counter-example. If the failure of reductionism and necessity of holism are supposed to be empirical claims about the biological realm—i.e., derived from contemporary knowledge about variation in biological processes like development—then they ought to be so for machines as well. We too often forget that machines are also empirical phenomena (Rapp 1981; Ch.1). Instead of beginning with a fundamental difference, echoing Ian Hacking[9] (1998, p.208) we should ask, "well, have we in fact made any holistic machines yet?" The claim that machines fail to display the right sort of holistic and emergent properties cannot be settled by examining only Descartes' favourite examples—clocks and

[7] Which were found eating the ivy vines around a garden in Cambridge.

[8] http://www.cam.ac.uk/research/news/functioning-mechanical-gears-seen-in-nature-for-the-first-time

[9] "Even questions such as 'Can machines think?' might be answered, in the spirit of his writing, 'Well, have we in fact made any thinking machines yet?' We have to ponder the matter, but we should not start from a fundamental opposition between machine and organism." [ibid]

waterworks—since there are other sorts of machines. The remainder of this section is devoted to arguing that there are indeed machines that are holistic and that even some familiar machines are likely included in that lot.

5.2.2 Holistic Machines

Consider the positive view of machines: that they are exhaustively dissectible into entities and activities that can be studied in isolation. This would not be a very interesting MOD if it were just a claim about the inefficacy of dissection and anatomy in organisms. That MOD would seem to turn on what surgical skills we happen to have at present, and not on any metaphysical aspects of new-mechanism. Rather, it is best seen as a claim about the absence of higher-order, system-level properties in machines: studying their particular constituent entities is sufficient to understand their wholes (unlike, supposedly, organisms with their "multiply-realizable developmental modules"). Machines are totally assembled from entities and activities that were designed in isolation, one might read. But this is false. There are cases of machines that were not so assembled and whose design did not proceed isolated-part by isolated-part. I discuss these first.

Lewens (2013) points to cases from "Evolutionary Electronics" where a microchip design is obtained by rounds of randomly generated configurations until a working whole is obtained. Interestingly: "some sections of the best-performing chip couldn't be altered without loss of function, even though they were not connected (in the usual way, at least) to the output of the chip" ([ibid] p. 645). In such cases, entities emerge whose activities cannot be studied in isolation, since their system-level effect (on the performance of the chip) depends on interactions between unintentionally and distally connected components. Moreover, for this holistic phenomenon to also be multiply-realizable would require only that some equally-well performing chips have different configurations. We may never see such behavior in clockwork, but that's no worse for the holism of machines generally.

Even outside an evolutionary design context the construction of a machine can result in holistic contributions of entities to system-level activities—features that were not designed in isolation. This need not be argued by appeal to any special positive characteristic of the design process, but can be seen as arising from the limitations of human intellect and intervention on our ability to prevent the emergence of that kind of holism. Holistic interactions can arise naturally in machines due to our ignorance. It is not without reason that Jacob (1977) and Jacob and Bendall (1983) referred to evolution as "tinkering": in a workshop the design process is less than ideal and can produce unexpected system-level effects from even small and unintentional modifications. There may be holistic interactions that arise accidentally in machines even when we do our best to design and assemble them part-by-part.

Considering the evolved chip again, we might reject this source of holistic phenomena in machines on the grounds that those effects were created by artificial selection, by imitation of organisms. But just as in organisms, artificial selection cannot create alone. As Godfrey-Smith (2014 p.41) notes, "it is only the combination of selection and mutation that is creative". The selection of chips with better performance takes place after the creation of chips, some of which already have beneficial non-connected system-level effects. Those effects themselves are the result of a process analogous to mutation—the random creation of chips with different configurations of components. What the artificial selection process does do is increase the likelihood of those random beneficial effects appearing in the final chip produced.

If the artificial selection process itself does not create non-connected effects, then we have reason to believe they appear in familiar machines as well, albeit at an unknown base frequency. As Lewens (2013) points out, "innovators may have unwittingly failed to reduce forms of cross-talk that, unknown to them, make an important contribution to the overall operation of the artefact". Non-connected parts can have a positive effect on the function of a machine, and since they emerged in that context merely from the diverse arrangements of electric components, they may also appear outside the evolutionary design context whenever diverse configurations do, e.g. the microcircuitry of common household teachnologies like radios and computers.

That fact is all that is required for a counterexample to the non-holistic view of machines. Overall contributions of unintentionally modified parts to the operation of a machine are holistic activities of constitutive entities to the higher-order phenomena of machine activity. Although the ordinary design process makes "exhaustive dissection" and "study in isolation" more likely—since it is the rational intent of the designer—it cannot ensure non-holistic behavior in the resultant machine. The new-mechanist rejection of the organism qua machine on the grounds of a poverty of holism is empirically unjustified, since system-level effects are not unique to biological systems. That of course does not mean that we should return to any sort of naïve mechanism—an analysis of organisms qua entities that are literally or functionally identical to machines (more below, § 5.3.1)—it just means that the sort of holism wanted as justification for new-mechanism is not going to suffice as an MOD.

If the MOD provided by Austin's (2016) new-mechanism is one that relies on the reducibility and dissectibility of static machines contra the emergence of dynamic and multiply-realizable organisms, cases like those involved in evolutionary electronics remind that there are or could be evolved or unintentionally holistic machines. The holistic-MODs in support of new-mechanism fail, just as the stochastic-MODs against it do. Therefore, if the metaphysical framework provided by new-mechanism is sufficient to analyze both sorts of phenomena then a new-mechanist as such would not need to defend an MOD. Austin

may be right about the sociological causes of processualist critiques of mechanism, but both processualists (Ch.4) and Austin have advocated MODs unnecessarily. That said, if machines can be holistic, new-mechanism will need some other way to distinguish itself from old-mechanism, since "organism qua (holistic) machine" is a not unattractive source of justificatory analogies in biology.

5.3 Mechanism Has Never Been *Machine*-Reductionism

Some retrospection at this point is helpful. The debates that have concerned us in the preceding sections (§ 1-2) and chapter (Ch.4) are not *sui generis*, but simply the most recent developments in a tradition of debates. The present cases of New-Mechanist vs. Critic, or Processualist vs. New-Mechanist, etc., have a long history. Most recently, these debates find antecedents in debates between Mechanists, in a broader sense, and their critics, Organicists or Vitalists. Moreover, we can trace them back at least as far as debates about the mechanical philosophy of Descartes.[10] While there is perhaps no single persisting debate at the core of this generation-spanning opposition between "Mechanists" and their critics or opposers, there is a persistent theme: the contentious role of machines in metaphysics of science.

It is thus tempting to believe, with the advent of the modern versions of these views, that this contention has been resolved.[11] That is, it is tempting to think that our modern grown-up new-mechanist metaphysics has overcome the vices of its antecedents, specifically the vice of equating machines with organisms. Since I have (above and earlier) argued that both new mechanism and processualism do not depend on MODs—despite how some have advocated or criticized them—I am at risk of contributing to this misinterpretation. In this final section I hope to set this straight: our new metaphysics have not outgrown the MODs they once depended on, since they never did depend on MODs.

5.3.1 Naive-Mechanism: A Persistent but Idle Straw-Man

It is both insufficient and naïve to characterize Mechanism as the view that being mechanical is a matter of being functionally or causally identical to machines. Let's call the view that mechanism is simply a claim about being identical to machines "naïve-mechanism". I call it 'naïve' not simply as a pejorative, but because it seems to be the view that we often consider first, and because 'native' or 'nascent' would come with other unintended connotations. Adopting naïve-mechanism is a vice on par with committing the *virtus dormativa* and

[10] See Riskin (2016).

[11] Perhaps exemplified by the processualism of Dupré and new-mechanism of Machamer et al. or Bechtel.

survives, in the public consciousness at least, in claims like "the body is a machine". It can only be half of the picture of mechanistic explanation in science, since it is only a non-trivial claim when made about non-machines. Clearly, naïve-mechanism vacuously explains machines—machines are mechanical because they are identical to machines. It may say something non-trivial about organisms that they are (more or less) analogous to machines, but then we must flesh this out by giving an account of machines and the properties they have that are ripe for analogy.

That said, naïve-mechanism is a far more idle view than is perhaps presumed by critics of Mechanism. It is a favourite straw-man; though not harmless. It sits in the background during contemporary discussions of Mechanism. For example, when Hacking (1998) praises Canguilhem for recognizing "how central to Descartes was the idea that animals are machines" (p.203), the naïve view, indeed, the naivety of the view and the apparent implication for the naivety of Descartes is background to the following joke,

> Chomsky (1962) found in Descartes just the man to be his predecessor, and used the title Cartesian Linguistics. So that distinction [between humans and animals, and thus machines] is very present to recent thinking about language. The more general doctrine, that animals are machines, does not rate high in today's consciousness. We think of it as vaguely quaint. It summons up images of Descartes kicking dogs downstairs.—Hacking (1998) p.203

This is not a slip specific to Hacking. It is common to attribute the origin of Mechanism as a metaphysics to Descartes' *Treatise on Man*, with its corresponding "mechanical philosophy" seen as an ontology—though there are historical antecedents who predate Descartes.[12] It is thus tempting to conclude that Descartes at least must have held to naïve-mechanism.

The problem is that this is not what Descartes says. After a very extended thought-experiment / description of the workings of a hypothetical machine, built by god, Descartes says explicitly what we are to take from this exercise.

> I desire that you consider that all the functions that I have attributed to this machine... and in this they imitate as perfectly as possible the movements of real men. I desire, I say, that you should consider that these functions follow in this machine simply from the disposition of the organs and *wholly naturally as* the movements of a clock or other automaton follow from the disposition of its counterweights and wheels. To explain these functions, then, it is not

[12] Rapp (1981) attributes the first instance of interpreting the "universe as a giant mechanical clock" to Nicholas Oresmus in the 14th century, but calls Descartes' mechanical philosophy the first "authoritative formulation" of such a view.

> necessary to conceive of any vegetative or sensitive soul, or any other principle of movement or life. —Descartes (1972) p.169, my emphasis

We are not told that real men (or animals) or their movements are those of machines, nor even that they even enjoy the same dispositions. We are told that a machine which imitates us in a manner as perfectly as god can create will have functions that follow from the dispositions of its organs. Since god can, presumably, create an exact copy of our bodies this implies that our functions also "follow from the dispositions of our organs". What we are given beyond this is a simile: our functions follow from the dispositions of our organs "wholly naturally as" the functions of a clock or other automaton follow from the dispositions of their parts. What is the same between us, or animals, and machines is not the functions nor the dispositions, but the naturalness of how functions arise from dispositions of parts, organs. Riskin (2016) notes the same tendency to misinterpret Descartes (embedding what I take to be an ontological point into the explanatory context of 'understanding') as follows,

> Descartes's proposal that an animal is a machine has sounded to most people... like saying that an animal is essentially inanimate... But it is a misreading of Descartes... By describing animals as automata, Descartes did not mean to reduce them to lifelessness. On the contrary, he meant to declare that one could explain every aspect of life in terms of machinery, and so could understand the workings of living beings *as fully as* a clockmaker understands a clock. —Riskin (2016) p.44-45, my emphasis

This is not an exercise in charitable Descartes-rereading. Descartes certainly gets a lot wrong.[13] Nonetheless, what he desires we take from his mechanical philosophy is not naïve-mechanism. What we are actually given in the mechanical philosophy of Descartes is something like reductionism, grounded in dispositions, where the analogy with machines serves the rhetorical role of highlighting an analogous case where (presumably) we are not inclined to imagine a non-reductive explanation in terms of "vegetative or sensitive souls".[14] Modernizing Descartes, we might rephrase by saying that organisms are reducible to their parts, like machines are. This is good rhetoric insofar most people are already reductionists about machines; "holistic" machines were not on Descartes' or his reader's radar.

We may of course still disagree with this mechanical philosophy, in at least two ways: (1) we may not believe that our functions arise only from the dispositions of our organs, or

[13] He is unaware of statistical mechanics and so fails to understand the movement of gasses and fluids (although one would imagine he could have concocted analogies were he to have known); the sorts of things he believes are included in the body are not, such as spirits in the blood and literal fire; and many of the processes he believes in, such as the heating of the blood in the heart, turn out wrong.

[14] How seriously we are to take the reference to 'dispositions' is hard to assess. Suffice it to say this debate does not likely turn on the correct metaphysical account of dispositions.

(2) we might not believe that the arising of our functions from the disposition of our organs happens as naturally as it does in artificial machines. That is, we might hold an MOD based on a rejection of reductionism, or find Descartes' analogy a poor justification. But this is plainly a very different sort of debate to the one invoked by naïve-mechanism. This is a trend. Mechanisms turn out to rely on similes with machines for justification, but are substantive claims of some other sort. The topic of the next section is how this manifests in Mechanisms advocated post-Descartes and pre-new-mechanism.

5.3.2 Mechanism*s* not Mechanism

It is somewhat misleading to refer to "new mechanism" as if it were univocal, as if it were a single theory. As seen above (§ 5.1), there are a variety of presentations of new mechanism that we are sometimes licensed to treat as unified just when we discuss features that are common among them. This multiplicity is magnified as we step back to earlier Mechanisms; in the beginning of the 20th century and earlier, there were a lot more theories going around with the name 'Mechanism' than there perhaps are called 'New Mechanism' today. This presents a problem of identifying what features are common to these diverse views that license sometimes talking about them as a unity, and thus a problem of determining whether there is any single MOD troubling them all. I argue that, if we are looking for a core or center-point around which to amalgamate or unify past Mechanisms, it is reductionism and not machine-analogies or metaphors that best serves this role.

What makes this reductionist core less than obvious is that their inclusion in Mechanism often comes with other substantive metaphysical claims. What those substantive claims turn out to be is varied. Some common themes for this extra bit include:

0 Dispositionalism (in Descartes)

1 Mathematical modeling

2 Materialist atomism or "corpuscularianism"

3 Narrow sense Mechanics, i.e. the study of movement

4 The scientific method

5 Determinism

These five aspects of Mechanism appear together in various combinations and with different proportions of emphasis. For instance, Needham (1926) sees mechanism as a claim about physical reductionism and the in principle efficacy of the scientific method.

> If all things in heaven and earth could really be thought of as capable of being fully and finally revealed to man by the scientific method ; if physics was not only the most fundamental science but also the most fundamental of philosophies ; if, indeed, there could exist no metaphysic but science, then of course the world would be safe for mechanistic biology, and a place fit for biochemists to live in. —Needham (1926)

For another example, consider Bennett's (1986) characterization of the state of natural philosophy in the 1600s. Though abiological, Bennett's characterization is typical of discussions of Mechanism.

> [T]he seventeenth century was a time of growing consensus in natural philosophy... based on an experimentally practiced mechanical or corpuscularian philosophy of nature... [Wherein] the mathematical qualities fully characterized ultimate particles or corpuscules, though there was no narrow commitment to a rigorous metaphysic [e.g. Cartesian], and sensible phenomena resulted from the purely mechanical interactions of particles—interactions that were in principle fully exampled by gross machines. —Bennett (1986) p.1

Here mechanism is practically equated with corpuscularianism. Mechanistic views are defined as those where (i) mathematical modeling applies to (ii) particles together with a reductionist account of how phenomena arise therefrom. Again, it is only after the view has been described that the analogy with machines is given: "mechanical" interactions are those that are "in principle fully exampled by gross machines". This is not a concession to naïve-mechanism: Bennett's aim is a history of the involvement of mechanics (people who make machines[15]) in the course of early science, so the assumption of analogy with "gross machines" is really emphasizing a part of (iv) the scientific methods of the time. In any case, this is a claim made about the physical analogy between machines and corpuscules, not organisms.

Mechanism and its criticism has too long a history to make this point by the multiplication of instances. For my purposes it is only necessary that we avoid making mistakes in assessing the relationship between mechanism and the MOD. To that end I will state what I take to be the two major vices affecting the debate, so we can avoid them.

The first vice is limiting the class of machines used for analogy to a narrow and familiar group, usually clocks and clock-like machines (Ch.1). "Clockwork" analogies and metaphors

[15]Speaking critically of how the history of "mechanical philosophy or corpuscularianism" has been parochialized by a focus on customary distinctions between "scholar and craftsman, or the intellectual and the mechanic" Bennett (1986) says that these terms have essentially biased the debate because "the 'mechanic' or 'craftsman' has been understood as someone unconcerned with higher science."

can be carelessly used by advocates of mechanism and are thus often the target of critics. This is a vice of parochialism: even if (when) comparisons to machines are important justifications for Mechanism, this should hardly turn on those features of machines instantiated only in 16th century clocks. If there is a deep connection to be found (denied) in the relationship between Mechanism and machines, it ought to be stable in the face of (at least minor) advances in engineering. Nonetheless, this vice is the easiest avoided. We need only broaden our comparison class—of machines, but also organisms—when placing weight on such comparisons. For example, if the aim is to criticize Mechanism via a disanalogy with clocks, *even if* the Mechanist takes this seriously (i.e., they do not point out at the real issue is about dispositions, or reduction, etc.), then they can always simply respond with an analogy to a non-clocklike machine.

The second vice is specific to critics of mechanism: presuming that Mechanism in general stands or falls with its most flawed characterization, here called "naïve-mechanism". This often takes the form of overestimating the place or significance of machine analogies in mechanism *qua* metaphysics (Austin 2016). In these arguments, the straw-men are clocks. The right view of this analogy, to my mind, is that it is entirely idle or dispensable unless it is serving as a justification for one or more of the substantial theses (i-v) typically associated with Mechanism. Criticizing this analogy in isolation thus both fails to get to the crux of the matter and to seriously address the substantial claims of Mechanists.

5.4 Conclusion: Poorly Defined Targets of Mechanistic Philosophy

In this chapter I have tried to demonstrate two things (i) MODs based on stochasticity and holism in support or criticism of mechanism have empirical counterexamples, and (ii) that despite the persistence of machine analogies and disanalogies in advocacy and criticism of Mechanism, these are dispensable or else false. My arguments for (i) should dispel remaining worries that rejection of MODs entail the adoption of a (potentially fraught) metaphysics. Argument for (ii) should go some way to explaining why the place of machines is so hotly contested in debates about the correct metaphysics of the biological sciences, and hopefully encourage some pause at the introduction of contingent facts about clocks into metaphysical debates about organisms. Both of these points, in their own ways, depend on recognizing that Machines–physical objects like clocks and waterworks—have been mistargeted in what is otherwise quite clearly a metaphysical debate.

A final word on this oddity in conclusion is helpful, before we leave metaphysics mostly behind in the subsequent chapters. Surely, if a theory claims to be metaphysical in the sense that it offers an ontology for the sciences, then it must *at very least* be empirically adequate. We should thus expect that any broad metaphysics of science, such as Processualism or Mechanism, should not turn on the analysis of machines or organisms, precisely because they are both empirical phenomena and thus things that these metaphysics should be adequate to analyse. Any MOD at the level of metaphysics would need to turn on whether a given metaphysics were adequate to one (machines) while inadequate to another (organisms), but this would, if true, be an argument against the metaphysics. Thankfully, in the cases examined in this chapter, this is not true: no MOD falls out of Mechanist metaphysics, since both machines and organisms *can* be analysed in Mechanistic terms.

Chapter 6

Evolution and the Machine Organism-Distinction

Abstract

Is evolution or evolvability the MOD? Organisms evolve, but do machines evolve; and if so, how? In this chapter I argue that an account of evolution broad enough to include all cases of biological evolution will admit cases of technological evolution. To support this, I examine two established accounts of biological evolution. The first, due to Lewontin (1970), defines evolution in terms of heritable variation in fitness. The second, advanced first by Dawkins (1976) and Hull (1980) and elaborated by Sterelny et al. (1996), defines evolution in terms of replicators and interactors. In both cases I show that there are technological changes that satisfy the definition of evolution. I then respond to a single counterargument: that machines fail to form the kinds of reproductive/replicative lineages required by both of these accounts of evolution. To respond to this counterargument I consider another account of the evolutionary process, due to Woese and Fox (1977) and Woese (1998), that defines evolution in terms that do not require lineages of organisms. If unclear reproductive lineage relationships are a problem for technological evolution, they are just as much of a problem for biological evolution. I take this to show that no MOD can be made on the basis of a sufficiently broad account of evolution.

6.1 Introduction

> [A]t an abstract level technology change can be described as "evolutionary". However... this fact does not entail that evolutionary approaches to technology

> will revolutionize the way we understand this domain. At present, there is no reason to expect the creation of robust, general models of technological evolution except at the most abstract, uninformative level.—Lewens (2006) p.140

What does it mean for technology to evolve? Since there are a number of accounts of biological evolution, there are a number of different things that it might mean for technologies to evolve. This chapter addresses three accounts of evolution: (i) evolution as individuals satisfying the Lewontin conditions, (ii) evolution as parts satisfying the conditions for being an (extended) replicator and, more unusually, (iii) evolution as collectives of entities satisfying conditions for being what Woese and Fox (1977) termed 'progenotes', what I call Woese conditions.

Perhaps most technology change does not occur by processes even — *pace* Lewens' charity — "abstractly" evolutionary. Though, sometimes it does, and it is those times that this chapter addresses. We may think those times so few or so abstract that they are not worthy of examination, and I present some cases to show that this is not so. However, to accept infrequency or abstractness as criticisms of evolutionary approaches to technology change would be a double-standard, since many biological evolutionary processes are also infrequent and abstract. Perhaps most biological changes are not evolutionary, and those that are are always at some more-or-less high level of abstraction. Speciation is quite infrequent and is abstract enough to cover manifold very different mechanisms, but knowledge about speciation is essential to understanding biology. It is enough to reject an evolutionary MOD that some (infrequent, abstract) cases of technology change are evolutionary.

The more serious problem with providing an evolutionary account of technology is the relationship between technological evolutionism and traditional theories of technological change. It is often argued that fitting instances of cultural change to evolutionary frameworks does not explain much, if anything, besides what is explained on rational, historical, anthropological, and/or sociological grounds (see Lewens 2015). However, rejecting an MOD based on the applicability of evolution to machines does not require the evolutionary framework to explain much besides what can be explained in other disciplinary frameworks. It only requires determining whether evolutionary frameworks are in fact appropriate to the phenomenon to which they are applied. In a rebuttal of an article[1] defending cultural evolution, Fracchia and Lewontin (2005) make a perennial point about the success vs. appropriateness of evolutionary explanations outside biological systems.

> The real issue... is not whether explanations can be successfully manufactured on the basis of paradigmatic assumptions, but whether the paradigmatic assumptions

[1] Runciman, W. G. (2005). Culture does evolve. History and Theory, 44(1), 1-13.

> are appropriate to the object of analysis. The selectionist paradigm requires the reduction of society and culture to inheritance systems... with society and culture thus reduced to inheritance systems, history can be reduced to "evolution." But these reductions, which are required by the selectionist paradigm, exclude much that is essential to a satisfactory historical explanation.—Fracchia and Lewontin (2005) p.14

I am not concerned here with what Fracchia and Lewontin (2005) call the 'selectionist paradigm', within biology or outside it, and leave questions about society and culture to others. I am concerned whether "paradigmatic assumptions" of various theories of technological evolution—those due to Lewontin, but also others—are "appropriate to" machines, and whether they exclude much that would be required of successful explanations. If theorists of cultural evolution have, in response to what they see as a lack of evolutionary explanations of technology, been exclusionary of other approaches, that is their error (Lewens 2015). But it is a rectifiable error. When evolutionary frameworks genuinely apply to objects of inquiry—whether machine or organism—they need not exclude the application of other theories. Evolution never supplanted physiology, ecology, etc. Technological evolutionism should likewise not supplant social sciences or history of technology. Disciplinary territorial dispute may be inevitable, but in-principle exclusion of non-evolutionary theories by evolutionary theory can be avoided.

This penultimate chapter only claims that evolution will not establish an MOD, since three of our best general accounts of evolution sometimes apply to machines. The first two accounts of evolution addressed here are familiar, the last is less so. In § 6.2 I examine the case for the evolution of machines conceived as individuals satisfying Lewontin's (1970) conditions for evolution by natural selection. There I argue that there is a subset of machines that fairly straightforwardly possess heritable variation in fitness. In § 6.3 I examine the case for evolution of machines conceived as replicators satisfying Sterelny et al.'s (1996) extended replicator framework. In the final § 6.4 I take up a more heterodox view of evolution advocated by Carl Woese, proposed to apply to the time before the divergence of the major kingdoms of life. This theory explicitly drops features thought key to defining evolution of present organisms by the other accounts. Significantly, it drops the requirement for organismal lineages. Woese considers the evolution of '*progenotes*'—instead of organisms, true *genotes*.[2] There I argue that even if lineage-formation is a major problem for technological evolution, it is also a problem for biological evolution.

[2]From pro + genote, an entity yet to possess a modern genotype-phenotype connection. Sometimes incorrectly assumed to derive from 'progenitor' and 'prokaryote'.

6.2 Evolution of Machines as individuals

This section argues that some machines, like some organisms, evolve as individuals due to their possession of heritable variation in fitness. Clarifying an analytic definition of evolution has been a "struggle for existence" among philosophers and biological theorists. The most famous condensation is due to Richard Lewontin (1970), as follows verbatim.

L1. Different individuals in a population have different morphologies, physiologies, and behaviors (phenotypic variation).

L2. Different phenotypes have different rates of survival and reproduction in different environments (differential fitness).

L3. There is a correlation between parents and offspring in the contribution of each to future generations (fitness is heritable).

Lewontin says that these are "three principles embody the principle of evolution by natural selection" [ibid p.1] (hereafter, Evolution by Natural Selection (ENS)). He then immediately asserts their sufficiency: "While they hold, a population will undergo evolutionary change" [ibid].

Does ENS in this sense apply to machines? This question requires some qualification. I think it is best qualified by taking a step back: What variable are we substituting for 'machines' when we check the applicability of the conditions? That is: Are machines supposed to be checked against L1-3 as "Different (Xs) individuals in a population..." or as "Different (Ys) phenotypes..."? If we were asking about the applicability of ENS to birds, we mean to substitute birds for the individuals that evolve (Xs), while if we are asking about the applicability of ENS to colouration we are asking about phenotypes (Ys), i.e. the sorts of things which may vary in a population of evolving individuals. The remainder of this section (§ 6.2.1-3) considers those machines to which ENS applies as individuals, the applicability of ENS to machines considered as (extended) phenotypes occupies the following section (§ 6.3).

6.2.1 Phenotypic Variation in Machines

ML1. Different machines in a population have different morphologies, physiologies, and behaviors (phenotypic variation).

Different machines are variable in their shapes, internal processes or mechanisms, actions and interactions. Moreover, they are so variable at present that they cannot be distinguished from organisms by any of the features thought fundamental to being an organism (Ch.2-5).

However, we might still worry that machines are barred from satisfying L1 if 'morphology', 'physiology', and 'behaviour' are given readings as specifically biological terms. I take the work of previous chapters to marshal against this worry, since it requires some MOD to define what is 'specifically biological'. That said, we can give independent reasons not to take L1 to be overly contingent on the biological connotation of 'phenotype' for organisms. This is because the Lewontin conditions are not thought to apply solely to organisms, but to other variable entities as well. Lewontin says that replacing 'individual' with 'population' and (re)interpreting phenotype in a "distributed" way suffices to make L1-3 applicable to the evolution of populations. L1-3 are, by proponents of MLS at least, thought to apply to genes, as well as species and even higher taxa (Okasha 2003; Godfrey-Smith 2009; Doolittle 2017). If genes or species, for instance, have 'morphologies' and 'physiologies', it is in quite a different sense than that for organisms. So we cannot bar an entity from having variation simply because that entity is a non-organism on Lewontin's account. Insofar as machines have 'morphologies' and 'physiologies', it is in a different sense from that of organisms. However, this is no reason to preclude machines from satisfying Lewontin's criterion of variation.

6.2.2 Differential Fitness of Machines

ML2. Different characteristics of machines have different rates of survival and reproduction in different environments (differential fitness).

Different machines persist for different lengths of time, and cannot help but do so on the basis of their particular characteristics. As discussed in Ch.4, this persistence is sometimes less "active" and more a matter of the durability of their components, but the same is true of organisms. Nonetheless, some machines also actively engage in processes that maintain them in the face of interactions with their surroundings. Computer viruses encode different processes that help evade detection and deletion by host software; computers contain anti-virus software that allows them to persist for different lengths of time in the presence of viral attacks. Likewise, examples abound of robotics—such as the Canadarm discussed in Ch.3—that are able to maintain themselves by replacing or repairing damaged components.

More questionable is the role of reproduction or replication in machines. In this section I am only concerned with showing that there are machines satisfying L2—not that machines generally or even often do—so it is only necessary to concentrate on the best cases of machine reproduction and the mechanical and digital self-reproducing automata are those best cases.[3] Perhaps the most interesting cases come from self-reproducing automata—robotic machines

[3]For a discussion of the more general problem of culture-level replicators see Lewens (2015).

that can copy themselves (see Ch. 3)—but the most frequently replicating are the digital, virtual, or *in silico* machines.

> *There appear to be programs that can reproduce... but which are not viruses. These 'computer organisms' may be a serious security problem...*
> *—Aldeman (1989)*

von Neumann began the work on self-reproducing robotics, but his ideas also spawned a number of computational self-reproducing entities. These were called "self-reproducing automata", and have taken many forms since their initial hypothetical introduction (see Stahl 1965). These computational automata, as well as later entities such as computer viruses, worms, and the "computer organisms"' mentioned above, are often attributed features of autonomy (Ch.3) such as self-maintenance and self-reproduction (Bedau 2003; Mange et al. 1996; Aldeman 1989 von Neumann and Burks 1966; Stahl 1965; Cohen 1987). Varieties of these autonomous capacities have been studied in computational systems since their introduction (see Penrose 1959), in both computer science and artificial-life research (see review Bedau 2003). These are perhaps the best cases of machines with appreciable rates of survival and reproduction.

Though not explicit in the L2, there is often a background assumption that reproduction of the sort required by ENS is *self*-reproduction. Considering Godfrey-Smith's (2004) typology of forms of reproduction, including "simple", "scaffolded" and "collective" varieties ([ibid], see Ch.3 § 3.4.1), the sorts of reproduction we see in machines is on the spectrum from simple to scaffolded reproduction, and arguably much closer to the scaffolded pole. Should this bar them from satisfying L2, from varying in their rates of reproduction? Arguably, no. Firstly, this typology of marginal forms of reproduction was developed for evolving biological entities, and there are plenty of evolving biological individuals that are scaffolded reproducers as well: genes and (biological) viruses most prominently. Moreover, some evolving organisms are also scaffolded reproducers, such as obligate parasites and other obligate forms of symbiosis. Secondly, argued previously (see Ch.3, § 3.4.2), there are examples of machines that are more simple, "self" or autonomous reproducers, and a growing number of theoretical examples on the horizon.

Less technologically sophisticated examples also include artefacts which are produced by consecutive rounds of copying. Perhaps all familiar machines are scaffolded allonomous reproducers, the survival and reproduction of which, in different environments, differs depending on both their make-up and scaffolding, both by ourselves and other machines (more in § 3-4 below). Ancient tools likely enjoyed short lineages of faithful copying; manuscripts were once copied by hand; arguably "memes" in the sense of digital media; and

generally whichever machines are simple enough to be produced by emulation from observed cases. All provide examples of technological objects that differ either in their persistence or how frequently they are reproduced.[4]

This is enough to show that there are machines with sufficiently autonomous or simple reproductive capacities to endorse ML2 in some cases (e.g. computer viruses and self-replicating automata) and, if we are inclined to admit other sorts of reproduction, that a scaffolded version of L2 also applies to both machines and organisms.

6.2.3 Fitness is Heritable in Machines

ML3. There is a correlation between parent and offspring machines in the contribution of each to future generations (fitness is heritable).

Lewens (2015, p.14) provides a reading of "technological replicators" as "items whose function it is to preserve resemblance between artifactual parents and their offspring". Satisfying L1-3 does not require replicators, though they are a plausible cause of individuals satisfying L3. In the context of L1-3, what we require is at least these "artifactual parent-offspring relationships", and those machines with marginally autonomous reproduction capacities, discussed above, are the better candidates. The question is just: When machines do reproduce, is there correlation between the parent and offspring?[5]

In silico machines with reproductive capacities tend to enjoy near-perfect heritability of their characteristics when compared to the imperfect physio-chemical mechanisms deployed for copying and error correction in biological systems—so much so that the creation of "evolvable" computer viruses often requires the design of systems that introduce variation (Iliopoulos et al. 2011), periodically violating heritability rather than maintaining intergenerational correlation. Likewise, when familiar machines are produced by rounds of copying there will be correlation between the product and the copied machine. More appropriately for

[4]This reproduction, moreover, cannot always be dismissed as explainable strictly in terms of our ends (Ch.2). Some machines will be reproduced whether or not they offer any intended benefit to the scaffolding system. For analogy, in a pair of canonical papers for *Nature*, Orgel and Crick (1980) and Doolittle and Sapienza (1980) described their notion of selfish DNA. They sought to explain both why there is so much DNA in genomes and why most of it seems to do nothing whatsoever for the organisms bearing it, i.e. why there is so much junk in the genome (Ohno 1972). Their explanation is that most of the DNA is or was selfish, being copied and degrading without any significant effect on the organism. Since *computational* viruses and worms likewise have no, or mildly detrimental, effects on the systems harbouring them, the notion of selfish *in silico* machines might likewise help explain the presence of digital junk. If we were to look across all computer systems effected by malicious *in silico* machines, we would likely find that "traces of the virus were left on the system" (Cohen 1987) following unintended replication events.

[5]I imagine I am not alone in finding Lewontin's phrasing of L3 odd. To my mind the essential bit is the correlation between parent and offspring, and expectations of contribution to future generations stems from this. I concentrate on L3 with this reading in mind.

the sophisticated and mass manufactured technologies, the clearest cases of intergenerational correlation will occur when copying of individual machines proceeds by production on the basis of a plan (design or schematic) that is itself copied. Here there remains some question whether it is the plans or the machines that satisfy L3. However, the Lewontin conditions do not specify the mechanism of correlation between parent and offspring.

***

What does this say about the significance of Lewonton's treatments of evolution for machines generally? Although there is a class of machines satisfying ML1-3, this class is restricted. That said, the number and rate of evolving entities might be underestimated, simply from considering the proportion of machines satisfying ML1-3. Someone wanting to understand the history of technology who was unwilling to grant the efficacy of technological changes wrought by machines satisfying ML1-3 would be missing a significant part of that history (e.g. every change caused by *in silico* evolution). The class of Lewontinian changes might not explain a significant proportion of technology changes, though it arguably still explains a very large absolute number of instances of technology change.

The deeper problem is what we should say about cases where the causal histories of the production and change of technologies do not occur in neatly separated parent-offspring intergenerational relationships, when the parent-offspring relations are many-to-one or many-to-many. This question is addressed more fully in § 6.4, where I argue that the massively reticulated nature of the historical processes of production and change of machines may justify a different, non-Lewontinian, account of the evolutionary process, but does not establish an MOD. Before that, the following section takes a different recipe for ENS and argues that it also is satisfied by a number of machines, though suffers from the same problem of entangles causal histories.

6.3 Evolution of Machines as Parts

Instead of Lewontin's conditions, this section considers accounts of evolution stemming from Hull's (1980) replicator-interactor framework (see [ibid] p.318). Hull's (1980) account proceeds by the definition of two types of entities (replicators and interactors) then defines selection in terms of them, as follows,

> Replicator: an entity that passes on its structure directly in replication.
> Interactor: an entity that directly interacts as a cohesive whole with its environment in such a way that replication is differential.

> ...[The] selection process itself can be defined:
> Selection: A process in which the differential extinction and proliferation of interactors cause the differential perpetuation of the replicators that produced them.—Hull (1980) p.318

[U]sing Dawkins' own style of teleological argument one could claim that the bird is the nest's way of making another nest.
—Bateson (1978)

Hull's ideas have already been applied to artefacts. Dawkins' gene-vehicle distinction is nearly the replicator-interactor framework, and his 1982 *Extended Phenotype* spurred a debate about the place of animal-artefacts in gene-centrism or, more generally, replicator-centrism. Bateson's view, quoted page-right, is iconic of the strong conflicting intuitions in this arena. Bateson himself evidently treated it as a criticism, reductio, of Dawkins, Dawkins responded in kind, while Sterelny et al. (1996) later took it on as a positive consequence for the analysis of animal artefacts. This section argues, in effect, that sometimes we are a machine's way of making another machine. In a Batesonian reversal, or Saturnalia of the neo-Kantian view of organisms and machines (Ch.2-3), we can shift our focus to machines and treat organisms as a means to their replication.

Bateson's criticism' is that Dawkins' selfish gene is merely a shift of perspective: the role of genes (replicators) in the creation of present and future phenotypes is one possible focal point, but not a privileged one, thus not justifying treating genes as a special and unique factor in development. At least, not justifying that focus to the exclusion of all else. We could likewise privilege our focus on the role of a phenotype, such as the behavioural consequence of nest-building, in the creation of future phenotypes. Sterelny et al. (1996) largely accept this criticism of the privileged role of genes and take a middle way: genes are special, since they are replicators, but not unique, since other things besides are also replicators. They call this view the Extended Replicator Framework (ERF). I elaborate this view briefly before showing its relevance to machines.

The ERF has the consequence that some animal artefacts are replicators. Bateson's claim—changed from a modus tollens to modus ponens—is correct, "perhaps always" (Sterelny et al. 1996, p.398) or for some nests and some birds at least. On the ERF, when an animal artefact is a replicator, it is an "active replicator", since "their properties... influence the probability of being reproduced" [ibid]. Such artefacts "form lineages" via their builders or makers, who "carry information through which the nest is replicated" [ibid]. The ERF is also etiological, requiring that such artefacts be "adapted" [ibid], sufficiently linked for "cumulative selection" [ibid, p.391], and that it must be the case that "the explanation for their existence and nature

is that earlier copies played a similar role" [ibid, p.393]. These are all claims about a subset of non-human animal artefacts. Nonetheless, they apply to some human artefacts as well.

Sterelny et al. (1996) aim to distinguish their view from the "holism run amok" of DST, so are quick to point to cases of animal artefacts that are not replicators. Most importantly are cases where an artefact does not increase the probability of artefacts (of the same sort) appearing in the future—no matter how significant that artefact may be in the lifecycle of the organism. Garbage cans are vital to opossums, though do not affect their future existence via opossums; "possum behaviour does not result in the net flow of new cans" [ibid, p.392]. Likewise, despite their evident biological significance, hermit crab shells are not replicators "precisely because the hermit crab is unable to influence the availability of a critical resource in the next generation" [ibid, p.397]. Since at times the unavailability of such a resource can drive extinction, Sterelny et al. (1996) refer to this as "sad evolutionary dynamics".

Genes and some non-human animal artefacts are special, since replicators, but not unique, since *some human artefacts are also replicators*. Importantly, human artefacts are able to influence their availability in the next generation—we make trash cans, perhaps especially the ones we like, and to maintain a flow of shells we farm snails. If we go extinct for artificial reasons, it is far more likely to occur precisely because we are so very able to influence the availability of things that are, for now at least, resources. Our sad evolutionary dynamic is wrought by our lack of forethought, not ineffectuality. Opossums might not be able to influence the availability of garbage cans but, in Bateson's form: sometimes a human being is a garbage can's way of making another garbage can. Complex machines are perhaps especially plausible extended replicators. They are the sort of things that are unlikely to arise *ab initio*, instead owing their existence in large part to the existence of earlier "copies" playing similar roles. Their properties influence their probability of being reproduced; and we carry information, in a very literal way, by which such machines are replicated.

In the remainder of this section I address possible objections to applying the ERF to machines: (a) whether the process of replication of machines is too complex for there to be lineages of machines and, relatedly, (b) whether the sort of replication process for machines is fundamentally different from the case of animal artefacts.

I address (a) more fully in Ch.7 when discussing Dawkins' *Extended Phenotype Theory* (EPT). Dawkins enjoins us to think of the difference between the causal pathways leading to conventional phenotypes and those to extended phenotypes as being a difference in degree of complexity. In Ch.7 I argue that we can generalize on this approach to view all artefacts, human and non-human, as extended parts of the organisms that are causally responsible for them. What we have on that view is a great diversity of artefacts varying in the complexity of the causal processes leading to their re-existence. But we do not obtain a fundamental

difference between, e.g., nest-making and clock-making. Both sorts of artefacts can be seen as (complexly generated) parts of organisms and thereby extended phenotypic parts of organismal-lineages. On this view, complexity of the process of replication is just not relevant to lineage formation. Machine replication is of an incredibly allonomous sort (Ch.3, see also § 4 below), perhaps highly scaffolded reproduction (Godfrey-Smith 2014), or complex and "indirect" copying. But that is not problematic on biological grounds. Though Sterelny et al. (1996) agree in principle (quote below), a sustained defence of this view will need to wait for Ch.7.

> We do not see complexity as problematic; for an e-mail copy of this paper is very indirectly produced, and depends essentially on many elements additional to the Word document... Nor need we suppose that genes, nor any other replicator, are "self-replicating".—Sterelny et al. (1996) p.396

On the second point (b), even if complexity itself is not a problem, the sort of complexity may be. One sort of complexity of replicative causal chains pertains to the level of autonomy compared to allonomy, the involvement of other entities. In the language defined in Chapter 3, there is complex variation in allonomous self-directed capacities of artefact production. Thus we might think that it is not complexity alone, but the allonomous sort of complexity introduced into machines that makes the ERF inapplicable to them. This same problem will also arise in connection with the EPT in Ch.7. However, complexity of the sort involving others in "indirect" causal chains does not prevent something from being a copy, either in the case of genes or other replicated entities. Moreover, most animal cases—beavers and termites especially—do not enjoy an autonomous production of extended phenotype. In most cases of the EPT, the causal processes leading to a conventional versus extended phenotype differ in complexity and differ in allonomy, so allonomous causal chains do not distinguish the production of machines from that of phenotypes. No particular beaver, only families of beavers, are a lodge's way of making another lodge. Perhaps no particular human, but only groups of humans, are a clock's way of making another clock.

To my mind, the problem for machines and non-human animal artefacts is application of accounts of evolution for individuals in cases where the more appropriate account is about parts. Lewens (2004) raises a related objection that human artefact evolution does not proceed by reproductive chains of artefacts—which is appropriate provided we interpret these as chains of individuals.

> The problem of applying the evolutionary model to mass-produced artefacts becomes apparent when we notice that successive "generations" of artefacts typically do not give rise to each other through chains of reproduction, but

> instead owe their production to a common cause. A batch of [cars] in 2003 is not produced by cars from 2002: rather, both batches of cars come from the same production line. —Lewens (2006) p.142

A bird's nest in 2002 is not produced by a nest in 2001. But most parts of organisms are not produced by those parts in earlier generations. A nose in generation F_1 is not produced by a nose in F_0; successive samplings of proteins do not arise from one another through chains of reproduction but owe their production to a common cause. Individual cytochrome-450 molecules in one organism are not produced by those in its parent; they are not even produced by those present earlier in the same organism.

Furthermore, one might think that the applicability of the ERF to machines would thereby supply an evolutionary MOD, since organisms generally are not phenotypes, not parts of organisms. However, *some* organisms are best seen as parts of evolving lienages, rather than whole lineages themselves. Social insects, for instance, do not give rise to each other through chains of reproduction, but instead owe their production to the queen. This tells us that being a member of a reproductive chain is not necessary or sufficient for evolving in either organisms or machines. Sometimes merely being a part of a reproductive chain suffices, and both machines and organisms occasionally are such parts (Ch.7).

***

The difficulty is not with seeing machines as parts of some reproductive or persisting chain of wholes but with identifying precisely what they are parts of. A Kantian ornithologist would stress reciprocal causation with artefacts (Ch.2, Ch.7.4.1), that nests are simultaneously common causes of birds and commonly caused by birds. If we think of machines as animal artefacts, as parts of a whole "the manifold elements of which are related to each other reciprocally as end and means" (Kant 2000), then we had better be able to identify that whole. It is much easier to claim that machines are sometimes extended replicators than to identify the interactors that they produce, those cohesive wholes that ensure replication is differential.

This is because the history of technology change is highly entangled, with specific machines sometimes cohering better with different wholes at different times. Put another way, to evidence that selection is operative in technological changes, on Hull's (1980) account, it seems that we must be able to individuate the "proliferation of interactors", and the history of technology is not obviously organized into such things. Indeed, this is very similar to the deeper problem with applying the Lewontin conditions to machines: the causal histories of the production and change of technologies do not occur in neatly separated parent-offspring intergenerational relationships. For both accounts then, there seems to be a problem with identifying the historical pattern of descent: that it is too tangled, too reticulated to support

ENS. However, in § 6.4, I argue that although this is generally a problem for applying these evolutionary accounts to reticulate entities, it is not an MOD. This is because there is reticulation in the histories of both organisms and machines, and moreover, there are good ways of providing an account of the evolutionary process even in these cases.

6.4 Evolution of Machines as Collectives

This section argues that the theory of progenotes, conceived primarily by Woese as a theory of the evolution of early life, is appropriate to the present evolution of machines. I argue that progenote evolution is distinct from the accounts of evolutionary change given by the Lewontin conditions (§ 6.2), in Hull or the ERF (§ 6.3), and is helpful in understanding the evolution of present technologies. I begin by showing that the problem of highly reticulated histories is already a problem in biological evolution. This is especially so in evolutionary microbiology and genomics, where community interdependence and horizontal gene transfer (HGT) tie non-autonomous organisms together into groups with scrambled genealogical trees. I then describe Woese's view (§ 6.4.1), situating it with respect to the problem of reticulated genealogies. Woese's view of progenote evolution rejects the requirement for "organismal lineage", yet still suffices to explain forms of cumulative evolution. Different genealogies conflict, on Woese's view, because the evolving ancestors of modern life did not occur in organismal lineages, so there were no true genealogies. The problem is then explaining how evolution can take place without genealogies (§ 6.4.2). To do this, I reframe Woese's theory in terms similar in form to the recipe of the Lewontin conditions, as three Woese conditions. Finally, I show how some machines, conceived analogously as "pro-automata", evolve according to these Woese conditions (§ 6.4.3).

The scrambled history of cultural entities is also lodged as criticism of Dawkins' meme theory—that memes do not have clear ancestor-descendent relationships (Wilkins and Bourrat 2001) or that "if there are too many parents then there are no parents at all" (Godfrey-Smith 2012 p.2164)—and the same might be said about technologies. Gould likewise emphasized the difference between biological and cultural evolution in terms of their different topologies, arguing that branching trees are appropriate to biological evolution (except in certain marginalizable cases) while reticulation is the norm for culture (Gould 1988; Gray et al. 2007). I have argued above that some machines form lineages to a degree, though the most lineage-like cases are not very common (*in silico* entities, § 6.1) and the most common cases are not the most lineage-like (tools as replicators, § 6.2). These are the best cases where technologies, like many cultural items, "stand in recognizable genealogical relationships" (Lewens 2020, sec.12). That said, even these best cases suffer from problems

with applying accounts of evolution; even among the most lineage-like technologies their genealogies are highly reticulated, deviating significantly from a tree-like pattern of descent with modification.

However, this does not mark a *fundamental* difference with biology, since much of the history of biology is reticulated as well: viruses, HGT, endosymbiosis and hybridization all have a reticulating effect on life (see Doolittle and Brunet 2016). Happily, there are good methods of handling reticulation in both domains; this problem has been known to both cultural and biological theorists for some time. Even before the problem of HGT was taken up significantly in biology (see Quammen 2018), Deleuze and Guattari (1988) were claiming that strictly tree-like genealogical methods in social science were unable to capture historical patterns, claiming that "transversal communications between different lines scramble the genealogical trees" (p.11). They based this conclusion on their reading of the work of French biologists on an hypothetical transfer of an ancient viral gene from rodents to domestic cats (Benveniste et al. 1975; Benveniste and Todaro 1974). Though, the recommended solution for both Benveniste and colleagues (1974;1975) and for Deleuze and Guattari (1988) are roughly the same: make use of reticulated phylogenies instead. Among others, Gray et al. (2007) and my prior work in Brunet (2016) show in some detail how the methods from phylogenetics can be used to study highly reticulated non-biological phenomena.[6] The right move, then, is not to reject evolutionary models of technology change on account of reticulation but to demand that reticulation be accounted for within those models.

O'Malley (2014) has cautioned against basing our account of evolution narrowly on examples of macroscopic organisms—charismatic megafauna—when the vast majority of living things are microbiological. O'Malley's caution is Wittgensteinian (PI.593): we should not nourish our thinking with a one-sidedly macroscopic diet of examples. Certain, often background, assumptions that hold true of macrobes break down in microbiology and genomics, leading to misunderstanding of the majority evolutionary process. Importantly for us, the interplay of horizontal and vertical inheritance plays only a minor role in the evolution of macrobes, while it plays a major role in the evolution of microbes.

There are macroscopic cases of reticulation. Slime moulds manage to form fruiting bodies in highly complex and adaptive ways, despite the fact that the colonial aggregations required come from millions of parents with at times significant genetic diversity (Hehmeyer 2019; Sathe et al. 2010). Slime mould fruits do not have clear ancestor-descendent relationships or, if they do, they are of the sort that have very "many parents". There are occasional HGT events in metazoans, particularly in plants (Richardson and Palmer 2007). And there are also occasional tree-violating hybridization events between species, even in animals, as Gould

[6]See also network methods / conception in Doolittle and Bapteste (2007).

(1988) acknowledged.[7] But the clearest and largest abundance of counter-cases come from microbiology and genomics.

Recent work by Koonin and collaborators (Wolf and Koonin 2013; Koonin 2015; 2016) argues that, over the timescales of the evolution of major bacterial groups, gain and loss of genes, rather than evolution of vertically inherited genes, predominates the evolutionary process. In one sense, considering individual asexual microbes as wholes, these always have a single parent, single grandparent (excluding forms of "bacterial sex" such as plasmid sharing). But considered at the lower-level of genomic parts, the number and unrelatedness of ancestors of a given microbe can be high enough to affect heritability over the timespan of diversification of bacterial groups. Not only do organisms and machines have a partially reticulated history, the processes of reticulation seem to have had a very significant effect. This is enough to show the failure of the MOD based on the reticulated topology of the history of technologies: the history of organisms is likewise reticulated.

In the remainder of this chapter I hope to set aside one remaining worry about evolution as an MOD. If reticulation is such a problem for the general accounts of evolution we have examined, and technology change is *only ever* so reticulated that these accounts are inapplicable, then what sense of evolution, what account, would remain for machines? To address this worry, I will examine the account of an earlier period of biological evolution offered by Carl Woese (Woese 1998; Woese and Fox 1977; see Doolittle and Brown 1994; Koonin 2014). Carl Woese offered a theory of the early evolution of life where "Lateral gene transfer... would not only contribute significantly to but also would completely dominate the primitive evolutionary dynamic" (1998). An appreciation of the interplay between horizontal and vertical inheritance lead Woese to a very different account of this "primitive evolutionary dynamic" which can inform our conception of the present evolutionary dynamic of technology (and biology too) precisely because of how the Lewontin conditions fail in both cases. In the remainder of this section I describe this account, relate it to the Lewontin conditions, and argue that it is appropriate to the messy, entangled and reticulated genealogical relationships characteristic of technological change.

6.4.1 Evolution in the Progenote Era

Consider the following (hypothetical) Figure.6.1. Given some distribution of present traits and an assumed branching pattern, one can infer ancestral states on the basis of assumptions about the likelihood of the changes that gave rise to the present state. Given the above, it is

[7] For the effect this has on the tree-of-cells for hybridizing species, see Doolittle and Brunet (2016).

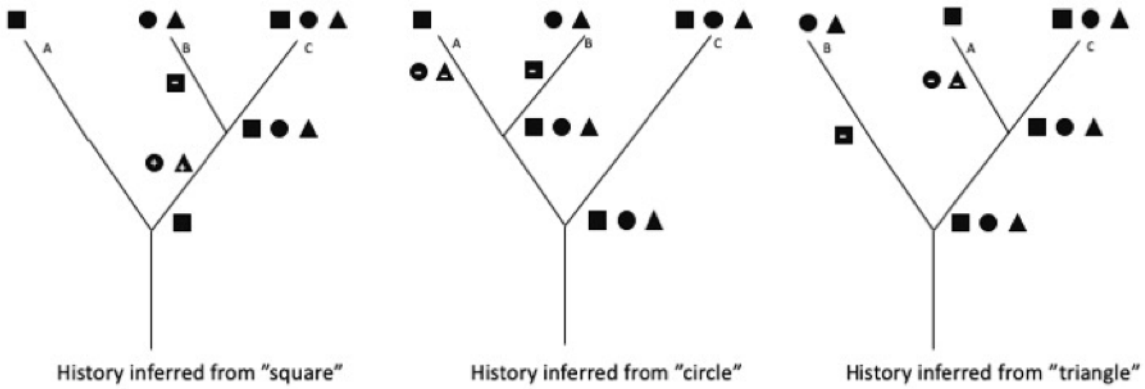


Fig. 6.1 Comparison of three hypothetical branching patterns of three lineages (A, B, C) with present traits (square, triangle, circle) and inferred ancestral relationships. Shapes at leaves indicate traits, shapes at internal nodes indicate best inferences of past traits, shapes to the left of branches indicate loss (-) or gain (+) of traits along branches.

entirely unclear which of the three should be favoured over the others, since nothing about the branching itself is favourable. Until recently, with the discovery of the Asgardarchaeota (see Williams et al. 2020), this was approximately the position we were in regarding the "tree" of life. The deep branching structure of the Bacteria, Archaea and Eukaryotes was uncertain. "Informational" 16S-RNA sequences indicated a tree where the Archaea and Eukaryotes were sister, while some gene phylogenies placed Eukaryotes sister to Bacteria, and some further placed Archaea sister to Bacteria. Without a reason to privilege one sort of evidence over another, the "true" branching pattern is underdetermined. Moreover, without an assumption about that pattern, inferences about the complement of traits that the Universal Ancestor possessed likewise remained uncertain.

These incongruities lead Carl Woese to an hypothesis about the origin of the major domains of life: the tree could be "rooted" not in a particular lineage, but in an amorphous conglomeration, one where the cellular and genetic lineages were uncoupled and the concept of genealogy, or organismal lineage, did not apply. On his view, the true branching pattern was not underdetermined by evidence, since there was no true branching pattern, only this amorphous conglomeration. He called these entities with decoupled cellular and genetic lineages 'progenotes', i.e., something that lacked but would give rise to an organism with a true genome.

So far 'progenote' has been characterized negatively: organismal lineages do not apply and they are not "genotes", but gave rise to them. In order to appreciate progenotes as more than retrodictive holism run amok, it we should see how Woese arrived at the notion, and provide a more concrete description. I first provide a short introduction to Woese's conception of the progenote, before turning to the consequences for our account of the

evolutionary process. For Woese, the progenote concept was a consequence of rejecting a "pivotal assumption" about early evolution, quoted in full below,

> **The Pivotal Assumption.** Most theories of early evolution tacitly assume that organismal lineages, organismal genealogies, have always existed and extend into the stage of evolution of the universal ancestor. Eukaryotes, of course, contain organellar genes, whose heritages are not those of the nuclear genes in general. Laterally transferred genes are seen in prokaryotes as well. Strictly speaking both eukaryotes and prokaryotes are of mixed heritage. Yet, we still speak of eukaryotic and prokaryotic "lineages" (and for good reason) because in both cases the vast majority of their genes presumably share a common history. If and only if this assumption holds, however, can we speak of organismal lineages and corresponding phylogenetic trees. But the assumption automatically makes the universal ancestor an organism that itself had a lineage, a discrete genealogy. The further back in evolutionary time we look, the more the notion of an "organismal lineage"—indeed, the very definition of "organism" itself—comes into question. It is time to release this notion of organismal lineages altogether and see where that leaves us.—Woese (1998) p.6855

Woese's identified pivotal assumption is that the vast majority of the genes in a cellular lineage share a common history. The progenote hypothesis comes when Woese rejects this assumption. The pivotal assumption is equivalent to assuming that inheritance is dominated by vertical inheritance and rejecting it is equivalent to asserting that vertical inheritance was not dominant, i.e., vertical inheritance was equal to horizontal inheritance or the latter dominated instead. There are many features of present life that prevent the predominance of horizontal inheritance. Thus the rejection of the pivotal assumption requires the rejection of some of the features of present life that ensure vertical inheritance dominates over horizontal. These are, (1) the tight association between the genotype and phenotype produced by an efficient and accurate transcription and translation system and, (2) physical features of modern cells, such as the existence of cell walls, sequestration of the germ line from environmental DNA, viral and endosymbiont defence systems, etc. Woese acknowledges the importance of (2) but focuses his attention on (1).

The modern translation apparatus is comparatively large, for a protein complex, in all known species. Woese and Fox (1977) explain,

> There can only be one reason for this [large size of the translation apparatus]; its size is essential to the accuracy with which the mechanism functions. There exists a direct correlation between the "size" of an automaton — as measured

> roughly by the number of components — and the accuracy of its functions.
> —Woese and Fox (1977) p.2

This is *prima facie* relevant to why there are no fully-fledged machine automata around today: our machines are composed of too few parts to do all of the things necessary to make all their components and put them together.[8] However, this is not obviously relevant to the issue of gene distributions and the prevalence of HGT—that requires two subsequent inferences.

> *Its [the progenote's] subsystems were generally less complex and hierarchically organized and the cell itself was less integrated than are cells today. The states of the cell were fewer, simpler and imprecisely defined and controlled.*
> —Woese (1998)

The first inference explains the connection to HGT by a size–accuracy feedback process, working backwards to infer an inaccurate and small ancestral state of the translation apparatus. Gains in accuracy entailed gains in the potential size of proteins translated, these platforming further gains in accuracy, and so on. Working backwards, earlier translation apparatuses were likely smaller and less efficient. But the translation apparatus is responsible for biosynthesis of all other cellular protein components—DNA copying, proofreading—so that the universal ancestor was likely limited in the complexity of other interactions. The next required inference is that entities with low integration are also likely to have high transferability of parts. The reason that (modern) amino-acyl tRNA synthetase genes are so frequently subject to HGT is, it is believed today, that they are quite "modular" or "not integrated"—they interact with only a few other cellular components. Induction on this leads to the conclusion that an entity that was less well integrated generally would also be more subject to HGT (see Novick and Doolittle 2020). Woese explains,

> The degree of connectedness of the componentry of the cell has profound evolutionary implications. If a cell was simple and highly modular in organization...all of the componentry of a cell could potentially be horizontally displaceable over time. The organismal genealogical record would be ephemeral; no stable record could exist. —Woese (2002) p.8744

Progenotes evolve, though by a mode that differs from that of modern organisms in significant ways. There is replication and reproduction of a sort, but these differ in their degree of precision. The evolutionary dynamic differs from that or modern organisms to such an extent that there is no genealogy. Or, if there is "genealogy" of a stripe, it is *not* the

[8] Though, of course, they do have some of those capacities (Ch.3) and most organisms cannot make all their components from scratch either.

genealogy of well-defined lineages presumed by our going accounts of ENS. Returning to the graphical representation of the problem presented above. The following (Figure.6.2) very roughly captures Woese's solution—an historical rhizome during the progenote era, with some lineage-like shoots to the present after the genotype-phenotype connection "crystalized" (see Koonin 2014). Deleuze and Guattari would have found Woese an attractive source of metaphors for their criticism of simplistic genealogy in social science.

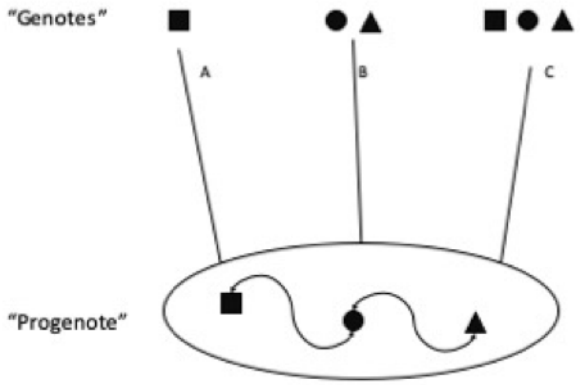


Fig. 6.2 Comparison of three hypothetical branching patterns of three lineages (A, B, C) with present traits (square, triangle, circle) and inferred progenote era.

If progenotes were so unlike modern exemplar organisms, what were they like? Woese intends the notion of progenote to be something like primitive cells, though in such a primitive form that it is not also an "organism". Woese and Fox are somewhat obscure about this. Woese and Fox (1977 p.3) say that "It is fair to say that the cell as we know it today would not exist", while Woese (1998) says "Evolution [in that era] can be seen as occurring on the subcellular level, although it actually happens in the context of (primitive) cells." Unfortunately, we are are not told precisely how primitive progenote "primitive cells" were. However, in his later view, Woese makes the positive claim that the non-organismal nature of progenote primitive cells is reason to treat only communities of progenotes as a unit of evolution. He writes,

> The ancestor cannot have been a particular organism, a single organismal lineage. It was communal, a loosely knit, diverse conglomeration of primitive cells that evolved as a unit, and eventually developed to a stage where it broke into several distinct communities, which in their turn became the three primary lines of descent...The universal ancestor was not an entity, not a thing. It is a process characteristic of a particular evolutionary stage.—Woese (1998)

During the progenote era of evolutionary history, large collectives, not individuals or their parts, are the appropriate "unit" for cumulative evolution. Presumably, the "communal" aspect of progenotes refers to their lack of autonomy as individual primitive cells, while referring to them as "loosely knit" or a "conglomeration" indicates that they are not specialized or particularly organized qua community—an ant hill is communal, but it is not loosely knit, and a herd of buffalo is loosely knit, but arguably not communal. For a perspective on the community of progenotes, Woese seems to have in mind something more than a mere group, but less than a society or super-organism.

> There are different ways of looking at such a community of progenotes. On the one hand, it could have been the loose-knit evolutionary (genetic) community just discussed. On the other, it could have been more like a modern bacterial consortium, with cells cross-feeding one another not only genetically but also metabolically.—Woese (1998) p.6856

Taking the other option, we can understand the communal nature of conglomerations of progenotes by analogy to contemporary community ecologies. In particular, the Black Queen Hypothesis (BQH) (Morris et al. 2005, or the GQH: Ch.3) helps us understand how interdependencies can exist at the level of collectives, binding them together into an evolving "unit". Consider two traits A and B, such that B is of some benefit (e.g. to the size of potentially translated proteins), only when supported by the effect of A (e.g. increase accuracy of translation). Usually, for there to be cumulative evolution of A+B in lineages dominated by vertical inheritance, those lineages must not lose A at a greater rate than they gain B (i.e. B emerges by mutation), otherwise they will not benefit from the synergy of A+B. But this assumption is violated by progenotes (and unstable ecological communities).

Nonetheless, it is still possible for there to be cumulative evolution of communal conglomerations with A+B, provided there is metabolic or genetic cross-feeding. Suppose the effects of A and B are non-local, like the "leaky functions" postulated in the BQH (Morris et al. 2005), as would obtain if A and B affect cross-feeding. Then, progenotes within a conglomations can benefit from the effects of A without physically having A within their primitive cell. Even if that progenote lineages loses A before gaining B, it can still benefit from the cumulative (non-local, leaky) effects of A, provided some other local progenote still contains A. Moreover, even if the effects of A and B are local (non-leaky), only having their cumulative effect when present in the same progenote, if there is "genetic cross-feeding" then a conglomerate with A and B is still more likely to give rise to A+B individuals and thus to enjoy the benefit of A+B. Even in present organisms HGT has been proposed to do this as an "adaptive strategy" for maintaining a beneficial trait within microbial communities. Even if you cannot hold onto your genes, if your community exchanges the benefits of

their pangenome or exchanges the genes themselves at sufficient frequency—if horizontal inheritance predominates—then cumulative evolution can occur.

This highly disordered, loosely knit genealogy arguably means that our general accounts of evolution do not apply well. We can see clearly why. The Lewontin conditions are best suited to cases of non-reticulated evolution or descent with modification, and apply less according as more reticulation obtains. They apply quite well when we assume that HGT is low. The Lewontin conditions do not apply well when HGT is too high since HGT is a constraint on heritability—if we obtain genes from our non-parents, this affects our resemblance to our parents. In the extreme, if HGT is high enough to overcome heritability, L3 will not be satisfied. Rephrasing Godfrey-Smith, if there are too many genetic ancestors contributing genes from outside a lineage then there is insufficient heritability within that lineage, correlation between parent and offspring lineages is washed away. We can take this as an argument against the evolutionary potential of entities subject to high HGT, or we can take it as an argument that the Lewontin conditions are just not the right account of evolution when HGT is high.

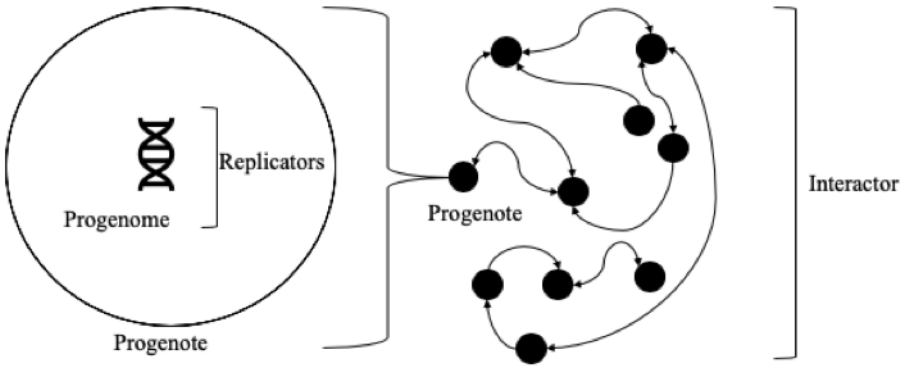


Fig. 6.3 Analogy between progenote framework and replicator-interactor framework. Over time, the effects of HGT on progenotes will decouple replicators from interactors, making the analogy only applicable over short timespans.

The replicator framework might be in a better position, but not by much. Progenotes do contain the ancestors of genes, replicators of an inaccurate sort, and conglomerations thereof do interact with the environment in such a way as to affect the differential reproduction of progenotes. The "unit of replication" would, on that analogy, be the ancestors of genes, and the "unit of interaction" or "basic unit of analysis" in such a framework will turn out to be the conglomerations or "cross-feeding consortia" of progenotes that Woese says "evolve as a unit" (Fig.6.3). That is approximately right at a given instant during the progenote era, although it begins to break down over the lengths of time of significant evolutionary change. Firstly, this

is because the protogenes are poor replicators, and secondly because progenote communities are poor interactors. Protogenes were inaccurately replicated and their ability to "code for" proteins was more stochastic. These entities were unlike what we would comfortably call genes today, and the same goes for pro-genomes as wholes. It is just not appropriate to take a gene's or genome's eye view in a non-genote. Moreover, the effect of HGT on decoupled replicators from interactors was supposed to be so significant that it dominated the evolutionary dynamic. Indeed, one of Dawkins' motivations for focusing of the gene as the evolving unit is, similarly, that organisms themselves are too ephemeral—especially sexual organisms—and progenote conglomerates are even more ephemeral.

The universal ancestor does have an evolutionary history, but that history is physical, not genealogical.
—Woese (1998)

Nonetheless, the inapplicability of these accounts of evolution to progenotes supports an account of technological evolution, since they fail to apply to this biological case for the same reasons they fail to apply to technology change: unclear or absent genealogies. At least as far as Woese's account is concerned, we do not need genealogy to have evolutionary history. Evidently the solution to the problem of patchy distribution of present traits leads to another: What should our account of evolution be, in non-genealogical entities so different from familiar organismal lineages? The Lewontin conditions wont help, since they require genealogies, making evolution in a progenote era forbidden by definition. I now turn to a reformulation of the Lewontin conditions in the spirit of Woese. I hope for this reformulation of Woese to help us understand progenotes in more familiar terms, but also for it to be directly relevant to our conception of technological evolution in the current era. In the following I provide an account of evolution that is appropriate to progenote conglomerations and show how it applies to some cases of technology change.

6.4.2 Lewontinizing Woese

Although progenotes are not paradigm organisms, they evolve as non-autonomous members of conglomerations. The notions of 'individual', 'population', 'phenotype' and 'parent and offspring', at least as we know them today and as they appear in Lewontin, are arguably inapplicable without significant qualification—qualifications that likely would, all things done, disqualify progenotes from satisfaction of those conditions. Nonetheless, they can be substituted for Woesian terms: 'progenote', 'conglomeration', 'physical characteristics', etc. However much the progenote era is unlike Lewontin's recipe for ENS, we gain some traction on the idea of progenote evolution by reframing it in similar terms, by separating out roles analogous to phenotypic variation, differential fitness and heritability. My attempt at such a Lewontin-style reading of Woese is as follows,

W1. Different progenotes in a conglomeration have different morphologies, physiologies, and behaviors (pro-phenotypic variation).

W2. Progenote characteristics affect the reoccurrence of their parts, via an effect on conglomerations, in an environment dependent way (mediated differential fitness).

[a.] Different characteristics of progenotes have different effects on different conglomerations (differential fitness-a).

[b.] Different characteristics of progenote conglomerations have different effects on the reoccurrence of progenote parts in different environments (differential fitness-b).

W3. There is a correlation between progenotes and those they physically influence in the contribution of each to future conglomerations (fitness is physically heritable).

A Woese-sympathetic Lewontin might have said: taken together these criteria are sufficient for evolution by natural selection in progenotes; any era during which these criteria are satisfied will undergo evolution among its progenotes.[9] Likewise, although individual progenotes are not auto- or self-reproductive, there can be selection at the level of conglomerations thereof provided they differentially contribute to conglomerations on the basis of their pro-phenotype. The need to separate Lewontin's second condition into W2.a-b is that progenotes themselves do not survive or reproduce when they directly interact with their environment; that they are loosely tangled together into "consortia" or "communal conglomerations". And W3 is Woese's notion of a history being "physical" but not genealogical framed in terms similar to those for heritability. Woese might have said, I imagine, that progenote conglomerations do have heritability, but it is "physical heritability".

6.4.3 Technological Progenotes

W1. Prophenotype in Machines

Let us call the technological analogue of the progenote the 'proautomaton' and our era the 'proautomaton era'.[10] Not all machines are (parts of) proautomatons. They must first of all

[9] Lewontin indeed says similar things to W1 when discussing the theory of Oparin (1957) on coacervates, which are membrane-like platforms for chemical reactions also hypothesized to be important in early evolution. Lewontin says, "Even though individual molecules in solution may not have been autocatalytic, there may have been selection among variants of a given molecular species when incorporated within a coacervate, so that the coacervate itself would evolve."

[10] For a machine to qualify as a proautomaton, my intuition is that they should have some of the properties of organisms or their parts. Proautomata should perhaps be "poietic" in the sense of the distinction between autopoietic and allopoietic offered by Maturana and Varela (1980); "organize" in the sense of the opposition between self-organizing and other-organized offered by Mossio and Moreno (2015); be teleological, due

not be too simple. At the same time, the technological equivalent of a progenote should not be as advanced as fully-fledged self-replicating machine-automata. Those, when or were they to exist, would be by hypothesis sufficiently like modern organisms to license the straightforward application of the Lewontin conditions. Automata are too much like "genotes" and simple machines are too simple, too much like parts (e.g. proteins) or extended replicators to count. We should probably include examples of all complex machines: computers, clockworks, televisions, 3d-printers, and perhaps tractors and jack-hammers. In every case, different sorts of complex machine have different properties—with potential for different effects on conglomerations thereof—and that is all that is required to potentially satisfy W1.

W2. Mediated Fitness Effects

Most machines at present are significantly allonomous; their effects on conglomerate entities will be mediated by ourselves and other machines. These mediated effects will be complex, but are no less evolutionarily significant simply for being mediated. Some complex machines satisfy W2.a when they have effects on conglomerations thereof, come whatever complexity of process. The more significant problem is nontrivial individuation of conglomerations for W2.b. This is not a special problem with proautomata or even progenotes—understanding the individuation of interdependent microbial communities is independently difficult (Inkpen et al. 2017). Though it does require posing the question in the right way: What are the higher level entities which are effected by complex machines and in turn affect their re-production?

Social sciences are replete with higher level objects of analysis—actor-networks, groups, programmes, corporations, societies, industries, institutions, etc.—many of which are, for better or worse, already the topics of cultural evolutionary theories. When the task is to identify conglomerations that reciprocally affect the re-production of machines then some of these higher-level entities might be good candidates. However, there is reason to worry that starting from these familiar entities and working backwards to infer the supportive effects on machines is liable to achieve little more than "show that a well-understood phenomenon can be reframed in an evolutionary idiom" (Lewens 2015, p.37). This would leave W2.a-b satisfied but uninteresting. Even admitting that machines can, at a certain level of abstraction, be thought of as "proautomata" and that the traditional candidates for higher-level entities as "conglomerations" thereof, there is reason to worry that this reading will be disfavoured over

to reciprocal causation with their conglomerations, in the Kantian sense offered by Nicholson (2013); sit somewhere on the autonomy-allonomy spectrum; and perhaps they should have sufficiently robust or active internal mechanisms or processes in the senses offered by Godfrey-Smith (2016), Skillings (2015), or Nicholson and Dupré (2018). Each of these features of organisms or their parts gets at something important about canonical examples of evolving things and, as each fails to be an MOD (Ch.1-5), both machines and organisms sit in this milieu.

the traditional modes of explanation available for each. *If* by individuating "conglomeratons of proautomata" we just end up picking out "industries", or similar, then arguably we have not individuated anything (special) at all. Fruitful application of an evolutionary account to technology change should provide something in addition to what we can already furnish with non-evolutionary reasoning about higher-level social entities.

That said, I don't see any reason that framing problems of technological change in terms of Woesian conditions should be attached to seeing conglomerations of proautomata as identical to any of these higher objects of analysis. And if it turns out that the higher-level entities that most clearly are affected by, and in turn affect, complex machines cross-cut traditional higher-level objects of analysis in social science, so much the better for the special explanatory status of the Woese conditions. If it turns out that it is only ever the traditional objects of social science analyses that turn out to satisfy the conditions for being a conglomerate of proautomata, then W2.a-b seem to be latching on to something independently interesting—insofar as those objects of analysis are interesting. Moreover, I do not see any reason that framing an instance of technology change in Woeseian terms should latch onto all the same features of those higher-level entities as are traditionally analysed. Institutions, industries, etc, are presumably interesting for a number of reasons besides their (W2.b) effects on the production of complex machines. However, an analysis of those higher-level entities qua proautomaton conglomerate should abstract away some of these details to supply a clearer picture of the evolutionary features of their dynamic—an evolutionary approach should cut out some of the gory details.

The aim of providing this account is not to show that it has the same theoretical explanatory power as those theories already used to explain technology change, but to "release" pre-existing notions of changing individuals—as Woese does with the notion of organismal lineage—and see where that leaves us. This demands considering conglomerations that become apparent when we relax our fixation on lineages, and thinking of complex machines as proautomata. Thinking of machines as proautomata, the aim is to consider the conglomeration that affects the re-production of a given machine. Recalling the problem we began with—that the histories of technologies, especially complex ones, are too reticulated to be accounted for with conventional phylogenies or genealogical trees—it is implausible that genealogies of any one of the typical objects of analysis alone can serve that role. The genealogical record obtained by looking too narrowly at a single sort of object of analysis will be too ephemeral.

This puts us in a position of trading tractability of analysis for plausibility of explanation—it is harder to pick out those conglomerations that do explain re-production (W2.b)—but this problem is not specific to proautomata conglomerations and obtains likewise for progenotes

and microbial communities. Perhaps it is better still to concentrate on those effects that seem to run counter to our expectations on more traditional analyses, cases where a "conflict" between the analyses reveals something important about the inter-level effects, e.g. when the effects that some machine has on conglomerations are not among those that are traditionally invoked to explain the existence or reoccurrence of the machine. For analogy, it is often remarked that MLS is only interesting when species selection pressures oppose those at the individual or genome level (see Jablonski 2008). Selection at different levels often aligns and when it does the lower (typically organismal) level is the favourite target of explanation. But when there is a conflict between levels of selection MLS becomes a more attractive explanatory framework. Likewise, cases of conflict between our traditional explanations of a machine and the effects it has qua proautomaton thereby speak in favour of considering it, and the entities it effects, within the Woese framework.

Personal computers (PCs) are a good example. PCs are highly designed objects. Nonetheless, they have many autonomous capacities and these can have effects on the persistence of conglomerations thereof. There are cases where these complex capacities align between lower and higher-levels of analysis, such as the positive effect of antiviral software on computer network stability. But there are also unexpected cases of conflict. To name one, PCs tend to engender dependence in users. Whatever effects some machine has that make conglomerations dependent on them—that make them "addictive", in the sense of "addiction molecules" (Van Melderin and De Bast 2009; Brunet and Doolittle 2018), or "selfish" broadly speaking—will tend to be re-produced so long as the conglomerate persists, regardless of whether this dependency was a reasoned product of design.

We might think specifically of designed cases like spell-check software affecting unassisted spelling activity or learning, or of more general forms of dependence wrought by converting one's livelihood to require PCs.[11] In cases like these it makes sense to apply a framework like W2 to understand the causal relationship between the effect of computation-dependence on the persistence of techno-industrial conglomerates (W2.a), and the consequent reciprocal effect on computer production (W2.b). To explain why there are so many recurrences of PCs, we will have to refer in part to the reciprocal effects of dependency, though that needn't ever have been one of the motivations of their users or designers, nor figure in economic accounts of their development, or be part of our account of the computer manufacturing industry conceived as an institution, etc.

Dairying also affords an interesting example, since it is a classic case of cultural evolution that has shifted drastically by the introduction of technology. Whatever role there may have

[11] Corporations, institutions or economists might have intended certain forms of dependence, which, as a precondition for demand, fits nicely into a traditional economic framework. But many such effects likely were not and could not have been intended.

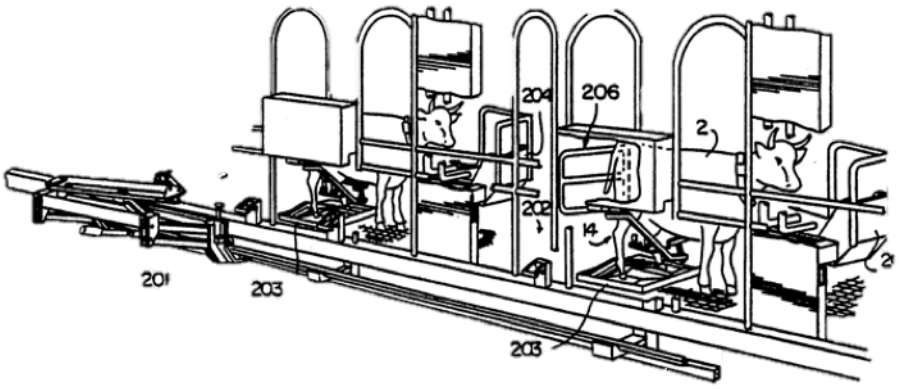


Fig. 6.4 Automatic milking machine. US Patent #: US5379721A

been for the effect of milk production to be brought about by the cultural evolution of techniques for milking by hand (Lewens 2015), these have evidently been largely supplanted by the technologies involved in production of automatic milking machines (Fig.6.4). Likewise, whatever artificial selection there was between milking by hand and the genes involved in udder development is now involved in a new coevolutionary dynamic, further involving the use of artificial intelligence in the technologies for detecting udder placement and timing of milking (US Patent # US7895972B2).

Straightforward explanations in terms of positive effects of intentions are more or less obvious, though of course not guaranteed to be true. Perhaps automatization of farming had and was intended to have a direct and positive effect on the persistence and recurrence of farms using automata, since automation is known to free up resources to be put to use elsewhere in the industry. But these are not the only sorts of effects on re-production of milking technology. For example, if the agro-industrial complex happens now to be dependent on these technologies, a return to hand milking would mean significant reductions in growth and the arguments of the above paragraph apply. Automatic milking machines might be addictive machines within any agro-industrial conglomerate, and might recur despite our best intentions to eliminate them from farming practices, since large farms without them inevitably fail.

Positive effects and reciprocal dependencies also interact. As Woese noted about the efficiency of automata generally: accuracy trades off against complexity. The complexity of autonomous machines also correlates with the number and variability of techniques of technological resources required to produce them. Perhaps automation technologies generally tend to be good examples of proautomata, since they tend to serve both a positive role, for

which they presumably were designed, and later generate dependency on the capacities freed-up by automation. But automation technologies are not easily produced, and generally depend on the prior existence of other such technologies—e.g. modern milking machines require microprocessors for their AIs. Thus we should expect that highly efficient automata of the sort that might have positive system-level effects would also be locked into a causal feedback loop with larger conglomerations of such automata.

No present proautomaton is self-reproducing. This is a problem for satisfying Lewontin's second condition. Though, it is precisely because proautomata recur only when they both affect and are affected by "communal, loosely knit, conglomerations" of proautomata, and thus only "evolve as a unit", that they are good targets for evolutionary explanation satisfying W2 instead.

W3. Conglomerate Heritability

Heritability is important for cumulative evolution. The "problem" that the progenote theory solves is that ancient HGT overcomes heritability. It does this by situating progenotes within larger conglomerations, effectively pushing the problem of the explanation of similarity-across-time, and adaptive feedback, up to a higher-level of organization. We must also recall that progenote conglomerations need not reproduce—they may of course split off into isolated parts, but that is not part of the explanation for evolution among progenotes. Indeed, once the conglomerate ancestor was able to separate into distinct groups, once lineages "crystalized" (Woese 1998, the "shoots" in Fig.6.2), the progenote era was effectively over. The appropriate notion of heritability applies during the progenote era at the level of conglomerates.

Early life and progenotes may have had too many parents to have parents at all—their status as genealogical entities is threatened by their reticulated origins—but a form of "heritability" sufficient for cumulative evolution obtains at higher-levels of organization nonetheless. I suggest the same is true of proautomatons. Complex machines do form short-lived inaccurate lineages—insofar as the term still holds meaning at such a threshold; they are sometimes allonomous or scaffolded reproducers in Godfrey-Smith's sense, or replicators in the sense of the ERF, or extended parts of organisms, in a sense to be explicated in the final chapter (Ch.7). But they are modular, so that even though the structure of one machine may have a direct causal influence on the structure of subsequent machines, that relationship is too weak and ephemeral to explain significant cumulative change along those short-lived lineages. There is little guarantee that any feature introduced into a machine (intentionally or otherwise) will be re-produced in its "descendent" machines, those that it causally influences. On the other hand, if a machine were not modular but highly integrated instead, this would place a strong constraint on copying of that machine. It would probably

be difficult to produce make-shift parts for a highly-integrated fully-fledged self-reproducing automaton, and so lineages of such machines would likely have a higher heritability.

Because proautomatons are highly modular, however, tracking their ephemeral lineages goes only a little ways towards explaining their features. Too many of the causes and effects of proautomatons reside outside of their genealogical relationships (Fig.6.5). At this level, feedback effects within a lineage will be slight and short-lived, so that lineage-level cumulative evolution is improbable, and intention or design end up on balance better explanations for any given feature. Nonetheless, these divergent effects of proautomatons do

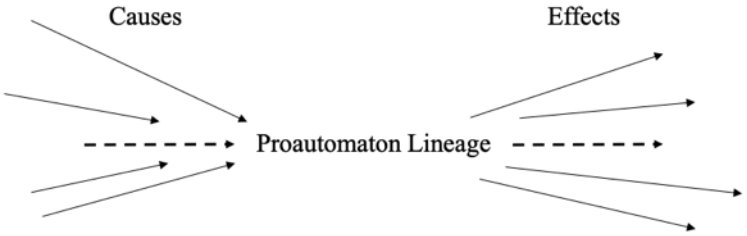


Fig. 6.5 Incident causes and divergent effects (solid lines) of proautomaton lineage (dashed line).

not exist in a vacuum; they ramify across the conglomerate, reticulate, "physical histories" of other machines (Figure 6.6 below). At the level of conglomerations, there is plenty of opportunity for complex forms of feedback—technological changes platforming further changes in other machines—leading to cumulative change at the level of conglomerates. So heritability in our proautomaton era should be assessed at this level.

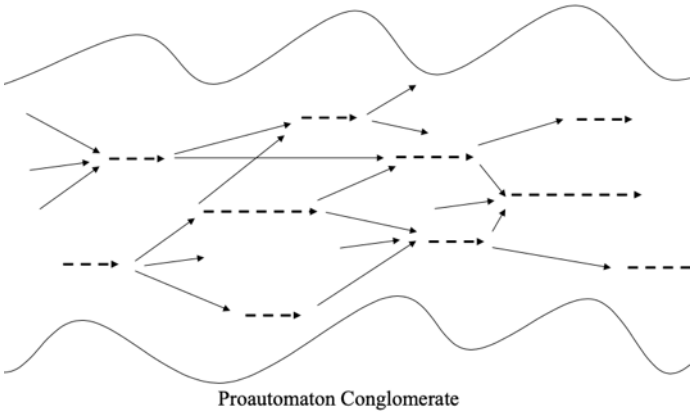


Fig. 6.6 Rendering of reticulated history of causes and effects (black lines) of progenotes / proautomata (dashed lines) within a conglomerate (wavy border).

Evidently a batch of milkers in 2002 does not contribute much to a batch in 2003—though they do contribute, perhaps vitally, to a complex system that ensures correlation among reproduced machines. Advances in computational power platform development of automation technologies that feedback to affect the production of computer hardware. Machines reap the benefits of developments elsewhere in the technological conglomerate, allowing the evolution of complex machine traits that would not have been possible within individual, inaccurate lineages of machines, or by intention alone. Contemporary technology change may not have a genealogical history sufficient for cumulative evolution along lineages of machines. However, as Woese realized about early life, this doesn't matter to evolution: physical history is enough.

6.5 Conclusion

I take the arguments of this chapter to show that no MOD based on evolvability will work. Organisms evolve sometimes as individuals, their parts evolve, and sometimes biological evolution takes place at the higher-level of larger collectives. So too for machines. Sometimes machines evolve as individuals. Though rarely, it does occur. Sometimes machines evolve as parts of other things, as artefacts as much tied to our evolution as nests are tied to the evolution of birds. That is, sometimes we ourselves are the "wholes" that machines evolve as parts of. The following chapter argues for this more fully. A relative significance debate is appropriate, and there may be no pressing need for engineers or anthropologists to change their good research programmes. However, evolutionary processes do sometimes occur in machines, so an MOD cannot be made on the basis of evolution or evolvability.

Chapter 7

A Machine-Organism Relationship: Part-Whole Classification

Abstract

In this final chapter I develop a view of the relationship between machines and organisms that is mereological: machines are extended parts of organisms. The notion of extension has received many prior treatments. Some version of this idea appears in Canguilhem (1952), Clynes and Kline (1960), Dawkins (1982), Haraway (1990), Sterelny et al. (1996), Clark and Chalmers (1998), Hacking (1998) and Kline (2009). What is common to these views is that they attempt to incorporate artefacts into some existing biological theory via part-whole relationships, as a functional biological object, such as an organ or phenotype. I adhere most closely to the terminology of Dawkins (1982), and propose that we consider machines as extended parts of organisms and of groups of organisms.

7.1 Introduction

> . . . [I]n artifacts we extend our own intentionality; our tools are extensions of our purposes... —Searle (1980) p.419

This final chapter is the most constructive. Instead of only countering MODs, I also argue for a particular relationship between some machines and some organisms. I argue that machines can be classified as biological entities, specifically, as the extended parts of organisms. Part-whole relationships are deeply connected with hierarchical classification in biology. Since machines are extended parts of biological things, they have a place in that classification.

Biological entities tend to be contained within larger biological entities.[1] The nature and ordering of that containment is a fundamental aspect of the classification of biological things. At the organism-level, we classify parts of organisms on the basis of their part-whole relationships. Typically, into a hierarchy of organelles, cells, tissues, organs and organ-systems. At levels above the organism, species-level and higher, we have come to see taxonomic classifications as metaphysically grounded in part-whole relationships between clades and metapopulation lineages (Hull 1986; Ghiselin 2005; De Queiroz 2007). I take this to be an example of real philosophical progress in biology: intuitive or similarity-based classifications of biological things can be replaced by classifications on the basis of part-whole relationships.[2]

The idea that machines can be included within our biological classification follows naturally from a solution to a problem that has already existed for some time: *the problematic position of animal-artefacts* (§ 7.2.1). A number of different thinkers (Weiner 1948; Canguilhem 1952; Clynes and Kline 1960; Dawkins 1982) have hit on this problem, in one way or another and for different reasons. Remarkably, they have come to very similar resolutions. Common to these resolutions is that animal artefacts should be seen as *extensions* of organisms. Whether parts of organisms or groups, animal-artefacts enjoy a place within the part-whole classification of biological entities.

This is more readily accepted in the case of non-human animal artefacts. Spider webs are extended parts of spiders; termite mounds are extended parts of termite colonies; birds' nests are extended parts of birds. But there is persistent suspicion that *human* artefacts cannot be lumped together with non-human ones. Dawkins (2004) is particularly critical of over-extending his analysis of animal artefacts to humans. I argue, to the contrary, that once we take account of the diversity of human and non-human animal artefacts, there is no sharp line separating the former and the latter. Put another way, a unified analysis of non-human animal artefacts spills over to include some human artefacts as well. If our biological mereology is expanded sufficiently for the classification of non-human animal artefacts, it thereby classifies some machines as well.

In § 7.2 I argue that animal artefacts and social organisms present a problem for biological classification, since they do not obviously fit into the part-whole hierarchy applicable to other biological entities, and that a solution to this problem is provided by the notion of extension. In § 7.3 examine the reasons given by Dawkins (2004) and Sterelny (2004) that the

[1] Aside from such maximally inclusive things as Life itself or the sum totality of the "tree" of life (Mariscal and Doolittle 2018).

[2] Many other questions in biology also have a mereological component. For example, whether or not some entity (e.g. a symbiont) evolves with another (e.g. a host) is often phrased as whether they are both parts of the same evolving lineage; key biological concepts (e.g. homology) are also plausibly interpreted as parthood relationships (Ghiselin 2005).

notion of extended phenotypes does not apply to all animal artefacts, and argue against these. In § 7.4 I present a general account of extended parts and show how it applies to include machines in biological classification. Finally, in § 7.5 I argue that this machine–organism relationship does not amount to an MOD, since some organisms are also extended parts of other organisms.

7.2 The Position of Non-Human Animal-Artefacts in Hierarchical Classification

7.2.1 The Problem

There is a clash between (a) the desire for a unified classification of biological entities and (b) problematic cases of biological entities that do not fit within the classification of traditional cases. These problematic cases include non-human animal artefacts, human artefacts, machines, and some organisms as well, especially when those organisms are social (bees), or artificial in some way (dairy cows, GMO mice). In her book *Animal Constructions and Technological Knowledge*, Shew (2017) develops a view of "technologies that considers some of the actions and products of non-human animals as technological in character." This chapter intends nearly the converse. I develop a view of animal artefacts that considers machines as biological in character. I consider this a development in its own right, the kind of view we should be aiming for after the preceding eulogy for the MOD. It is also a solution to a problem that pre-exists in biology. Around the organism-level of classification, there is a great diversity of problematic cases where it is difficult to decide whether we are dealing with an organism, an item or artefact in its environment, or rather an organ, or a group of organisms.

To isolate this problem for the present discussion, two theoretical considerations must be backgrounded. I will discuss there first before elaborating on the problematic position of animal artefacts. (1) I adopt a view of classification that is mereological (based on part-whole relationships). I take it as a genuine discovery of biology that higher-order taxa (above species) are constituted by part-whole relationships among lineages, and thus that phylogenetic classification *is* classification on the basis of part-whole relationships (Hull 1986; Ghiselin 2005; De Queroz 2007). Moreover, this mereological view of classification is in many ways continuous with classification of lower-level taxa (at species level and below). Groups are parts of species, and organisms themselves can be described in terms of part-whole relationships: organisms are composed of organ systems, which are composed of organs, tissues, cells, organelles, multimers, etc. (2) There are problems and incongruities with

the mereological classification system, most notably (i) the misalignment of mereological (AKA "organizational") levels with putative levels of selection, and (ii) the known processes of reticulation amongst biological parts that prevent the mereological classification system from being strictly hierarchical at all levels. The phylogenetic history of life contains many events that violate a strictly tree-like pattern of dissent, such as symbiosis, endosymbiosis and horizontal gene-transfer (Deleuze and Guittari 1988; Doolittle 2000; Doolittle and Baptest 2007; Doolittle and Brunet 2016). Indeed, unity *across* classification, levels of selection and folk intuitions about the desirability of hierarchical taxonomy may be precisely the sort of thing we should *not expect* from biology on metaphysical grounds. Life is too disorderly to expect such hierarchies (Dupre 1995). Nonetheless, I take it that both i-ii can be independently dealt with in satisfactory ways (Godfrey-Smith 2013, p.117; Brunet and Doolittle 2018; Doolittle and Brunet 2016; Doolittle and Bapteste 2007) and, moreover, that neither poses a particular problem for the present use of mereological classification—I see no problems with non-hierarchical mereological classifications that do not align with levels of selection.

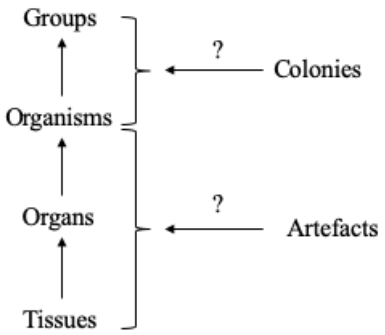


Fig. 7.1 Problematic place of artifacts and colonies within a portion of the biological hierarchy.

More threatening are cases where it seems that a biological entity is ambiguously classified or, worse, not classified at all. Two sorts of biological phenomena that lead to these problems are: animal artefacts and (eu)social relationships between animals (artefact animals). Consider a few examples. If a spider's web is an organ of the spider, it is the only such organ that is not composed of tissues; it is composed of macromolecules, but if it is an organelle of the spider it is the only such organelle not composing a cell. Worse, if a beaver's dam is a part of the beaver (lineage, life-cycle), it is not even principally composed of biomolecules originating in beaver bodies. Artefacts seem to stratify the biological hierarchy, spreading it out laterally in ways that forbid more conceptually-comforting nesting relationships among parts. For examples of the second sort we can consider social organisms, such as holobionts, chimeras, and crustaceans or insects with a caste system including non-reproductive members. Such social relationships cast doubt on a clear distinction between organ, organism and group, so that we might plausibly see individual worker ants or social shrimp, for instance, as

organisms in a group or as organs of an organism. In each of these diverse cases, the proper place of some entity within the hierarchy of biological organization is not clear (Figure 7.1).

Concentrating our focus around the organism-level, we can see that the problem is one of finding the right place for biological entities within the part-whole classification. Colonies, consisting of a number of organisms, seem to be groups. But qua parts of lineages, they seem to be organisms (super-organisms perhaps), which would make individuals within the colony more akin to the parts we usually call organs. When non-human animal artefacts consist of parts of organisms like proteins they seem to be organelles, but sitting outside cells seems to forbid this. Likewise, artefacts like tools and nests seem to be produced by organic activity and play such an integral role in animal life-cycles that they should be organs, despite not being composed of tissues in the right way. In short, our meso-organism-level mereology is a mess.

Further attention to specific examples will continue throughout the remainder. What is essential and problematic about these cases is now clear: part-whole classification at and around the organism-level seems unable to account for putatively biological entities—spider webs and worker ants. Some reformism of part-whole classification is needed to account for these phenomena. Unfortunately, there is no consensus as to how these problematic cases should be resolved. The following sections (§ 2.2-3) present some options and argue for a reform, not to our classification procedure itself, but to the nature of part-whole relationships admissible within it.

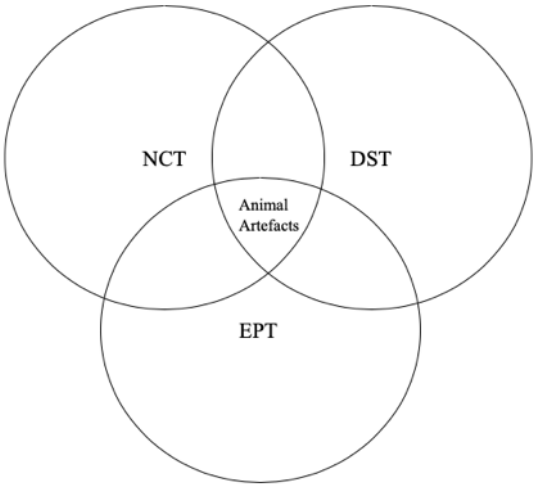
7.2.2 The Solution(s)

> Dawkins' extended phenotype analysis raises the option of treating animal artefacts as part of the organism, not part of its world. Not all artefacts fit into Dawkins' extended phenotype model.—Sterelny (2004) p.11

Perhaps the most famous account of artefacts in the philosophy of biology is provided by Dawkins' *Extended Phenotype*. On Dawkins' account (more details below) of non-human animal artefacts, some artefacts can be interpreted as *phenotypes* of the organisms that produce them. Thus we are licensed to conduct studies into the *genetic evolution of artefacts* (Dawkins 1982). Dawkins does not intend his view to impact our account of biological classification. Instead, his thesis that artefacts may be treated as phenotypic parts of organisms is justified on the basis of geneic selectionism. If we take gene selectionism (of Dawkins' selfish-gene variety) seriously we should see, his account proposes, that bodily or conventional phenotypes are not the only evolutionary relevant consequences of genes.

Artefacts too are consequences of genes, albeit sometimes at a much "longer reach" from the body.

There are alternatives, critics and contenders to Dawkins' Extended Phenotype Theory (EPT), each of which offering some perspective on artefacts and how they are best incorporated into biological theory. The two most noteworthy adversaries are the Developmental Systems Theory (DST) of Griffiths and Gray (2001; 2005) and the Niche Construction Theory (NCT) of, among many, Odling-Smee (1988; et al. 1996), Laland (2003), Jablonka (2007). Explicating the contentious relationship between these notions is beyond the scope of this chapter (but see Sterelny 2001). What each of these theories share is some proffered way of extending our analysis of organisms beyond the boundaries of their body and into the environment—the "world" in the epigraph from Sterelny (2004) above—to artefacts.



Wilkins and Bourrat (2001) say of the difference between Dawkins' EPT and the Extended Replicator Framework of Sterelny et al. (1996), the "views are so general that any case that can be described in one can be redescribed in the other. Differences lie in ease of description." The EPT, DST, and NCT are arguably also sufficiently general to permit redescriptions of the case at hand, and any one might be a viable route to a general theory of animal artefacts. However, in the remainder of this chapter I primarily discuss the EPT of Dawkins, because it furnishes us, without any redescription, with a notion of *extension*. The notion of extension is a general one: it has been offered a few times without the Dawkinsian undergirding, and largely before Dawkins' *Extended Phenotype*. Other theorists have used similar concepts of extension for similar uses: the analysis of human artefacts, tools, and in cybernetics (Weiner 1948; Canguilhem 1952; Clynes and Kline 1960; Haraway 1991; Hacking 1998). The EPT is just a better starting-point than the NCT or DST for incorporating these other non-Dawkinsian views.

Claims about treating parts as "extensions" have arisen from strong analogies between machines and *organs*. Hacking (1998) points to Canguilhem's (1952) *La Connaissance de la Vie*, where he advocates the view that machines are a kind of organ, an "organic projection". Hacking applies the term extension' to Canguilhem's phrase 'organic projection'. Early cybernetic discussions of automata by Wiener (1948) likewise draw parallels between organs and machines, though not explicitly as "extensions". Wiener at times sounds like Canguilhem, "...we deal with automata effectively coupled to the external world... The organs by which impressions are received are the equivalents of the human and animal sense organs" (Wiener 1948). Haraway (1991) points to a different strand beginning in Clynes and Kline's (1960) article *Cyborgs and Space*, a speculative discussion of the use of machines to create cyborgs, "exogenously extended" organisms [ibid p.1]. Along the strand stemming from both Dawkins and Clynes and Kline, Clark and Chalmers (1998) and Clark (2001) offer their own view of 'cyborgs' via the notion of extended cognition. It is not essential to cyborgs, on Clark's account, that any parts be internal to the organism, since some parts of the organism (specifically, cognitive resources) are already extended beyond that boundary. Objects like computers and notebooks can become the bearers of cognitive properties (e.g., beliefs) and thus parts of biological systems (Clark and Chalmers 1998), in virtue of how they extend cognition. See the chronology in the historical sketch below (Figure.7.9).

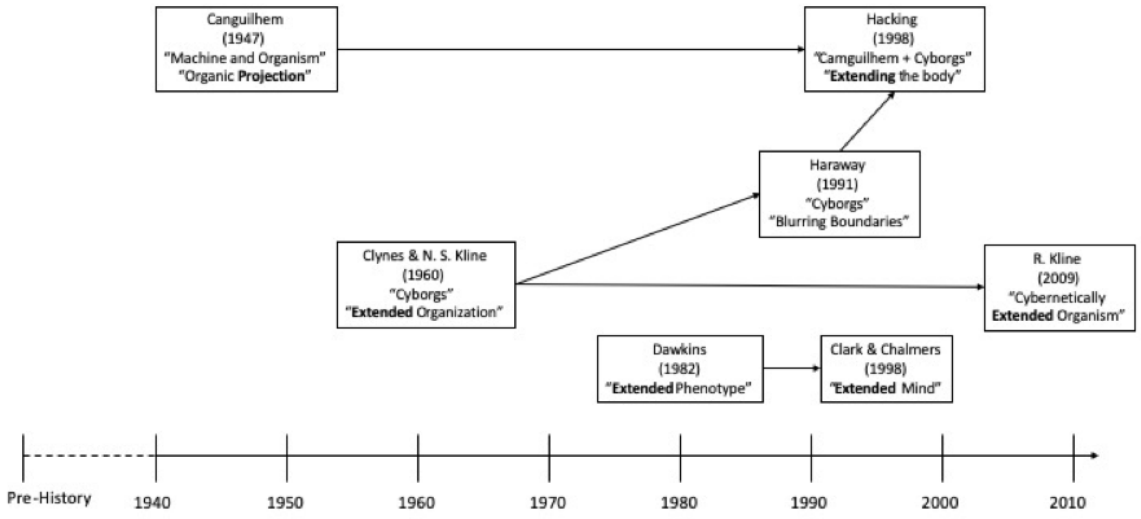


Fig. 7.2 Short history of concepts of technology, machines or animal artefacts as "extensions" or "projections" of organisms. Arrows indicate influence on subsequent articles.

The following quotes are representative.

> . . . [M]achines can be considered organs of the human species. A tool or a machine is an organ, and organs are tools or machines. —Canguilhem (1952)
>
> For the exogenously extended organizational complex, function as an integrated homeostatic system unconsciously, we propose the term "Cyborg". —Clynes and Kline (1960) p.1
>
> . . . Once we have accepted that there are genes for building behavior, the rules of existing terminology imply that the artefact itself should be treated as part of the phenotypic expression. —Dawkins (1982) p.302

Despite the different purposes of these authors, they are all interested in understanding animal artefacts from within their respective biological fields and all remarkably come to deploy similar concepts to that end. There does not seem to be any initial historical connection between them. Dawkins does not seem to have read or referenced Clynes and Kline, nor do they seem to have any engagement with Canguilhem. The word 'extension' and the nearly synonymous 'projection' have been used repeatedly and independently to refer to things that are in some sense parts of organisms but somehow more than the familiar stock of organs.[3] I interpret these authors as saying that biological mereology needs to supplement the stock of conventional parts with additional *extended parts*, and conclude from this that the proper mereology of biology is an extended hierarchy (Figure 7.3).

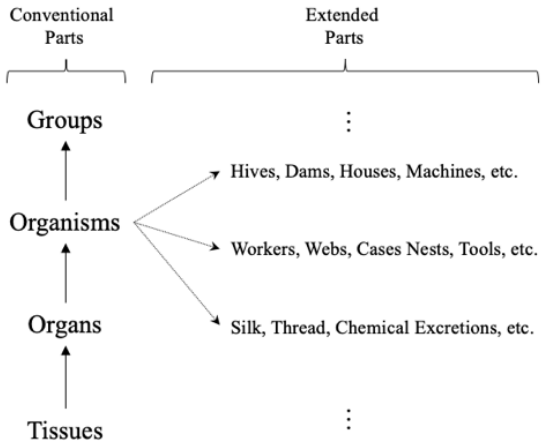


Fig. 7.3 Portion of the simple hierarchy of "conventional" parts (left) extended horizontally to include artefacts and colonial relationships as "extended parts" (right).

Why adopt this view of artefacts as extensions? There are as many reasons as theorists using the term, however, there are two main shared reasons. For some, the notion of extension

[3] A survey of the appearance of this idea in popular culture would be expansive.

is appealed to because it has the consequence that some previous distinctions are broken down. In Dawkins, Clark and Chalmers, the artefact/environment or artefact/phenotype distinction begins to be eroded. In Wiener, Canguilhem, Hacking and Haraway, the denial of the machine/organ distinction is taken to license an even broader "blurring" of the boundaries between the technological and biological, or natural and artificial. Hacking (1998) writes, "Canguilhem may have blurred more boundaries, in 1947, with specific argument and example, than Haraway did forty years later, by decree." To my mind, it is better not to adopt a notion of extension as a means to counter the MOD, or similar distinctions, but to adopt it because the MOD has already been eroded for other reasons.

Another shared reason for adopting a notion of extension is that it helps make sense of existing theory in bioscience, perhaps by carrying some theory to its "logical conclusion" (Laland 2004 p.314). In the case of those views stemming from Clynes and N. S. Kline (1960), R. Kline (2009) argues that it is the "key concepts of cybernetics" which lead naturally to the view that cyborgs are cybernetically *extended*. On Kline's (2009) view, we can only make sense of cyborgs as distinct from organisms, i.e. such that not every organism is trivially a "cyborg", when we understand them as cybernetically extended.

> At first thought, 'cybernetic organism' seems like a misnomer because all organisms are cybernetic in that they interact with the world through information and feedback control, [which are] the key concepts in cybernetics[.] The mouse and implanted pump is thus a cybernetically extended organism – an organism extended with cybernetic technology – what [Clynes and Kline] call a cyborg.
> —R. Kline (2009) p.332

For Dawkins' part, he says that it is the "rules of existing terminology" (Dawkins 1982) that demand we see some non-human animal artefacts as parts of the phenotype. If we are going to make sense of non-human animal artefacts in genetics at all, then they will have to be construed as (complex, indirect, behavioural) consequences of genes, and on existing terminology the consequences of genes are phenotypes. So artefacts are part of phenotypic expression. To distinguish artefacts from the other, conventional or bodily phenotypes, Dawkins uses the term 'extended'. However, Dawkins is resistant to further extension of his notion of the extended phentype (see Laland 2004; Dawkins 2004), and particularly to the inclusion of human artefacts as extended phenotypes. The following section argues that we can and should generalize the EPT to apply to human artefacts, to at least some machines.

7.3 Pushing the Limits of the Extended Phenotype Theory

Dawkins (2004) says that the EPT does not apply to human buildings. Sterelny (2004) also argues that the EPT does not apply to some simple tools, some non-human animal tools included. What sets these limits on the EPT? There seem to be two sorts of reasons given that some artefact should not qualify as an extended phenotype. In this section I explain and argue against these reasoned limits on the EPT, before offing a more general framework of extended parts in the following section.

The first relates to genetic adaptationism. Apparently, if an artefact is not produced by genes in such a way that the effects of the artefact, potentially at least, affect the frequency of those genes—if it is not a "candidate adaptation" of some genes—then it cannot count as an extended phenotype. The second limit on applying the EPT to machines relates to the level of specificity, plasticity, generality or flexibility involved in the production of the artefact. Apparently, if an artefact is produced by a process that is too general, flexible or plastic, then it cannot count as an extended phenotype. The two are entangled, since it is some specific set of genes that are required to be candidate adaptations.

Consider Dawkins' (2004) response to Laland (2004),

> The relevant point about the extended phenotype is that it is a disciplined extension. There are lots of other tempting 'extensions', which sound similar but take us off in misleading directions. I have always fought shy of misapplying the phrase to a profligate range of apparently plausible extensions.
> ...[W]hen I am asked...whether buildings count as extended phenotypes, I answer no, on *the grounds* that the success or failure of buildings does not affect the frequency of architects' genes in the gene pool. Extended phenotypes are worthy of the name *only if* they are candidate adaptations for the benefit of alleles responsible for variations in them...—Dawkins (2004) p.377, emphasis added.

We should reject part of Dawkins' claim straight away, before dealing with the more challenging claim about candidate adaptations. Dawkins' first point, *the grounds* above, is a claim about reciprocal fitness effects and is false. Plainly, buildings do sometimes affect the frequency of their architects' genes in the gene pool. In 2015, a pair of UK news agencies reported that the architect of the Pyongyang Airport was reportedly executed by Kim Jong-un, because the leader did not like the design.[4] This is not an isolated example. The success or failure of shelters has a profound effect on human life; poor shelters cause death by the

[4](Accessed Nov 2019): https://www.dailymail.co.uk/news/article-3143059/Was-architect-Pyongyang-Airport-executed-Kim-Jong-didn-t-like-design.html

elements, disease or predation. Moreover, if the grounds were only that artefacts needed to have potential fitness effects (on the genes) of artefact producers, then perhaps all human artefacts are thereby included. Fitness effects are cheap. Cellphones kill drivers; a compass can save your life; clocks keep you from missing your date. Reciprocal fitness effects on builders' (genes) will not exclude human artefacts.

Dawkins' conditional *only if* claim is the more serious objection. If something is an extended phenotype, then it must be a candidate adaptation for the benefit of alleles responsible for variations in itself. The adaptationist component requires some qualification. There are plenty of phenotypes that are not "adaptations for the benefit of" anything. Moreover, there is nothing preventing an organism from building artefacts that are deleterious. Chickens will sometimes make nests and attempt to raise a brood outdoors, although in fact they are much safer from predators if they use the artificial nest boxes provided in their coop. But the condition should not be about being an adaptation but simply about being a candidate for adaptation; it is about whether some artefacts are really *traits* or *phenotypes* after all. Perhaps buildings do not and have not had a significant enough effect on builder genes so that variations correlating with good (or bad) building behaviour have ever contributed to the frequency of those genes in the population sufficiently to overcome drift. In that case, buildings would not be adaptations, but they would still be *candidates* for adaptations if the gene's effect or correlation had been sufficient to positively affect gene frequencies.

Fig. 7.4 One of my family hens in nest-box. Photo provided by Kathryn Price.

The limit boils down to the genetic causation of the artefact, to whether an artefact is a consequence of alleles responsible for variations in it. Dawkins' phrase "candidate adaptation" is a bit obscure, Sterelny speaks more plainly: woodpecker tools are not an aspect of the bird's phenotype. Surely artefacts have reciprocal effects on their architect's genes, the issue is really whether they have any effect on genes specifically *for* the artefact. If this is really what was required for human-artefacts to be subject to an extended phenotype analysis—that, e.g., there be clock-genes and microwave-genes—then the theory's domain of applicability stops short of a general analysis of artefacts. There are no clock-genes, we

may safely assume, if what is meant by clock-genes are genes that *specifically code for* the construction of clocks. Clocks are not polypeptides.

However, Dawkins' condition is stated in terms more general than a simple requirement for "genes-for" an artefact. He says there must be "alleles responsible for variations" in the artefact. This only says that there must be covariation between genes and artefacts—some weight to the "genes for" relation—not how specific that covariation must be. This invites two further questions: (q.1) what the sort of covariation is required and (q.2) how specific does covariation need to be, i.e., what grain of analysis of covariation is acceptable. A behavioural-genetic account of animal artefacts would be significantly constrained in the analysis of real cases if it required gene-artefact covariation to be overly direct, or overly specific. Perhaps most animal artefacts are not biochemical products of genes (and behaviour), like spider webs, nor perhaps do they covary with small specific sets of genes.

For instance, among artefacts that are acquired "ready-made" from the environment—the spikes of butcher birds, stone hammers of primates, bark spades of Visayan pigs—there can be little to *no covariation* between the structure of the extension and the genes of the organism. Consider hermit-crabs in particular. Hermit crab shells are animal artefacts if

Fig. 7.5 Hermit crabs lining up to exchange shells following a period of growth, smallest to largest. Obtained from (Nov 2019): http://www.bbc.co.uk/earth/story/20141103-hermit-crab

anything is, though they are not biochemically, synthetically, products of hermit crab genes. Instead they are the result of chains of crab behaviour—perhaps, more akin to mobile cave-dwelling than architecture. Nonetheless, hermit crab shells can benefit alleles responsible for variations in hermit crab shells (the specific shell worn by a specific crab) since they can benefit crabs that have behaviours that allow for variable choice of shells—or trading variable

shells, as in the image above. While there is little to no crab-gene/shell-structure covariation, the shell remains a reliable animal artefact due to the strong covariation of crab-genes and crab-behaviour for shell acquisition.

Moreover, the inclusion of artefacts that are products of behaviour alone is just what Dawkins promises. Dawkins' argumentative strategy while introducing the notion extended phenotypes is to show how it can be justified in very little and "not difficult" steps. This begins by noting that "conventional" or "bodily" phenotypic traits are often caused in "long, ramified and indirect causal chains" [ibid, p.198], and asking us to take a small step further for cases of animal artefacts. These too are caused by long, ramified, causal chains, it is just that those chains also involve behaviours, such as building.

The upshot is that it should not be covariation between human genes and human artefact structure but covariation between human genes and the behaviour that result in humans having artefacts. Humans end up with artefacts in perhaps as many ways as animals do—we acquire them ready-made, we find them, we shape them to greater or lesser degrees, we construct them from raw materials, etc. But this diversity of means opens up *more rather than less* opportunity for covariation between genes and behaviours. Of course, there has never been a genome-wide association study of correlations between gene variants and human artefact related behaviours. Though, it is difficult to imagine that there would be none—especially pertaining to shelter and simple tool-use. If some covariation is all that is required, and "long, ramified and indirect causal chains" are allowed, the EPT should arguably apply to a much broader class of artefacts. Once attention is shifted, in Dawkinsian fashion, away from the structure or material composition of the extension and towards the behaviours that result in it, it becomes far less clear that humans do not display precisely the right sort of covariation between our genes and artefact-resulting behaviours.

What should we say if the *only* genes that covary with human artefact behaviour are those that also have some non-specific effect on learning, intelligence, sociality, dexterity, etc? Sterelny (2004) provides a related limit on the EPT: that it is curtailed by plasticity.[5]

> Treating artefacts as part of the agent's phenotype is not an option for artefacts whose construction is an expression of phenotypic plasticity: tool use in the woodpecker finch (the hypothesis runs) is not itself an adaptation, and is not an aspect of the bird's phenotype. The relevant phenotype trait is the learning ability through which the tool use is acquired.—Sterelny (2004) p.13

Sterelny (2004) also offers an explanation for why plasticity excludes some artefacts but not others.

[5] Laland (2004) offers a similar qualification based on the "flexibility" of learning.

> The tool of the woodpecker finch... is not likely to be a capacity that evolved incrementally through the assembly of dedicated developmental resources. In contrast, nests are aspects of a bird's extended phenotype. They are as developmentally stable, as heritable, and as predictable in their ecological effects as other traits. And they are adaptive complexes, and like other adaptive complexes, this nest almost certainly evolved incrementally. —Sterelny (2004) p.12

Since human artefacts are plastic to an upsetting degree, this would prevent the EPT from applying to them. However, plenty of conventional phenotypes did not evolve "incrementally through the assembly of dedicated developmental resources", so that shouldn't be part of our criteria for extended phenotypes in particular.[6] Plenty of phenotypes are irremediably tied up in complex epistatic causal chains that tie their expression, development, to a broad group of other genes. Dawkins' own style of argument should again be used to include these cases within the EPT. Many "conventional" or "bodily" phenotypic traits are often caused in "long, ramified and indirect causal chains" and some of these causal chains involve plasticity. Conventional, or bodily plasticity is no argument against being a phenotype.

By maintaining consistency with Dawkins' argumentative strategy we can push the limits of the EPT quite significantly towards including more artefacts. Then it is unclear why any human artefact should not be an extended phenotype as much as a bird's nest is. Moreover, we can keep the general notion of extension developed within the EPT while removing some of the commitments that limit its applicability to artefacts. To that end, the following section provides an account of *extended parts*, as a generalization of extended phenotypes.

7.4 Extended Parthood and Biological Classification

7.4.1 A General Account of Extended Parts

In § 7.2.2 I raised a number of accounts that deploy a similar notion of extension. They are a disunified bunch. Between Dawkins and Canguilhem there isn't a word in common, until Hacking provides his interpretation "projection = extension". I take Dawkins' EPT to be the best point of departure. However, as argued in § 7.3, Dawkins' own view of the limits of the EPT (Dawkins 2004) is inconsistent with the analytic strategy used to justify it (Dawkins 1982). Put another way, we have a theory of extended biological parthood in "many accents".

[6] As with Dawkins above, the physiological issue of how a given artefact is caused (by genes and during development) is run together with the status of the artefact as an adaptation. And as above, we should dismiss that part of the limit on the EPT that depends on the evolutionary status of the trait.

So which account of 'extension' should we adopt? In this section I offer a more generic, "accent free" characterization of the view to be used going forward.[7]

I begin with a definition of extended parthood and follow with examination of cases to show that it covers the previous uses of 'extension'.

> Part *p* is an extended part of *O* iff:
> **(1) Causality:** *O* or *O's* conspecifics contribute causally to the occurrence of *p*. E.g., *O* might produce, acquire, construct, synthesize, build, etc. as its contribution to *p*.
> **(2) Reliability:** *O* or *O's* conspecifics reliably contribute to the re-occurrence of p. E.g, *O* might reliably re-produce *p*, reliably re-acquire *p*, etc.
> **(3) Reciprocal Causality:** *p* must contribute to *O*. Minimally, *p* must have some reciprocal causal role in the life of *O*, and may further be classified accordingly as that effect was / is / is not under selection and had / has / does not have a positive, negative or neutral consequence on fitness.
> **(4) Detachment:** *p* must be a physically extended part of *O*. Part *p* must at least temporarily be detached from the conventional body of internal organs.

These criteria are intended to capture the general features of the EPT, without imposing the specific constraints (§ 7.3) that have come along with it. On (1) we would not want p to be a part of O, let alone an extended part, if O made no causal contribution to p. An organism, for instance, make some causal contribution to each of its conventional parts, and it should do the same for its extended parts. However, (1) does not build the features of genetic causation present in the EPT into this account of extension. Arguably one of the reasons that genetic causation of animal artefacts is preferred is that it entails some heritability or "developmental stability" (Sterelny 2004), and that is captured by (2), the requirement for reliable re-occurrence of the part. However, how reliability is ensured is left open by (2). On (3), as argued in Ch.2-3, there is something right about the idea that reciprocal causality is present in organic wholes, it is just wrong that these whole coincide with (the conventional bodies of) organisms—perhaps all of our conventional parts have reciprocal causal relations within us, but so too do our extended parts. Note that (3) does not build adaptation into our account of extension, so it makes sense to ask of some extended part whether it is also an adaptation. Finally, extension is meant to capture a parthood relationship that obtains even for (4) detached parts.

I now show that this account covers existing usage of 'extension'.

Consider a spider's web: it is (1) produced by the spider, (2) reliably re-produced by that spider and its descendents, (3) a contributor to the spider's fitness and (4) not one of the

[7] I owe the phrase "accent free" to the lectures of Peter Johnstone, Department of Mathematics, Cambridge.

spider's internal organs. Or consider a beaver's dam: it is also (1) produced and gathered by that beaver and/or its kin, (2) reliably re-produced by groups of beavers, (3) presumably not a luxury, and (4) not an internal body-part of any beaver. Likewise, when we think of Canguilhem's notion that human machines are "organs of our species" it is clear that a tool at least, such as a hammer (1) is produced by some human or group thereof, (2) is reliably re-produced by humans, and (3) contributes to human functions, such as domicile construction, and (4) is detached from the main portion of our organs. Taking an example from Clark and Chalmers (1998), consider an external cognitive resource such as a notebook: (1) it is produced by humans, (2) re-production of notebooks can be relied upon in many contexts, and (3) the ability to record information has effects on humans that can be positive, negative, or neutral, and again (4). Finally, the perennial example: a clock is (1) produced by individual humans or groups thereof, (2) re-produced (and repaired or sustained) reliably and (3) contributes to human functioning, such as by external "regulation" of circadian rhythms.

Turning to R. Kline, and Clynes and Kline's cyborgs, do cyborgs have extended parts? Considering some astronaut with an implant, e.g. an ordinary pacemaker, it (1) is surely produced by the astronaut or conspecifics, (2) can be re-produced as reliably as the conditions for its manufacture and implantation, and (3) hopefully contributes to, rather than detracts from, the functioning of the astronaut. On (4), though a pacemaker is internal at the time of its effects on the beating of the heart, it is exogenous, at least temporarily detached. Though, not all the exogenous parts of a cyborg are extended parts. For instance, note that a consequence of this definition (which I am willing to accept) is that Rose, the rat with attached osmotic pump, does not have that osmotic pump as an extended part. That is because Rose and her conspecifics do not causally contribute to the occurrence of the osmotic pump. For the same reason, the nest-box of my chicken in Fig.7.4 is not an extended part *of that chicken*. Of course, were Rose not around she could not be implanted, but nothing Rose or her conspecifics do makes the occurrence of her implants any more likely. The relevant causal difference makers are entirely human. This is acceptable for a number of reasons, not least of which is that it makes it clear that Rose was being mistreated. Moreover, it does not prevent Rose's implant or the nest-box from being an extended part of some human.

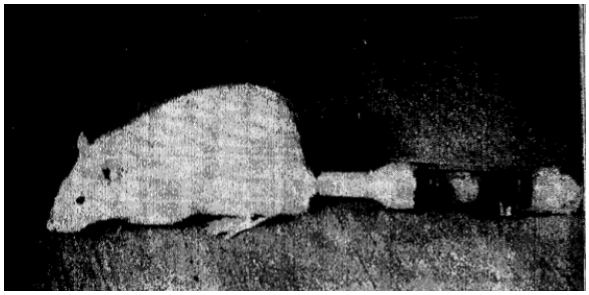


Fig. 7.6 "Rose", a rat implanted with an osmotic pump. Claimed by Clynes and Kline (1960) to be the "one of the first Cyborgs".

Works of art are also an example: they are certainly produced and reliably re-produced by us, although their reciprocal causal role is dubious. It is an open question what effects human art truly has on our lives as organisms, and whether human artistic abilities depend on a history of selection for such abilities, despite the enthusiasm of some (see Dutton 2009). Nonetheless, artistic behaviours are not evolutionary epiphenomena; they at least sometimes affect human lives reciprocally as paid crafts. There are also cases of uncontroversial (sexual) selection for the production of animal-artefacts that seem to occupy a middle-ground between watercolors and beaver dams, such as the nest of the bower bird (pictured below).[8]

Fig. 7.7 Bower-bird, rearranging some of the items it has collected to adorn its nest, top-right

From the foregoing examples it looks like quite a few biological, and some technological, things are extended parts. Is the definition too permissive? Are there things that count as artefacts that we might not want to consider our extended parts? Each criterion indicates at least some phenomena that should not be extended parts. E.g., (1) without a causal contribution to O, p should not be an extended part. All else being equal, a waterfall that provides drinking water is not an extended part (despite satisfying 3-4); a natural barrier such as a mountain-range or other happenstance bulwark is not an extended part; astronomic phenomena are not extended parts. A sextant is an artefact but stars aren't. That said, of course, an organism can have behaviours which make them seek out or otherwise secure the use of things that they otherwise have no causal contribution to. A bear seeks a cave; cliff-nesting birds squabble for rocky outcroppings that they could not sculpt. These latter cases satisfy (1) after all.

[8]Obtained from National Geographic, 2019. https://pmdvod.nationalgeographic.com/NG_Video/982/183/55695_1_1280x720_1024x576_177550915953.jpg

Likewise, (2) reliability varies. The happenstance arrangement of sticks fallen from a nest may satisfy (1,3-4)—they may be caused by nestbuilding and happen to confuse predators—without being explained by anything reliably occurring in bird behaviour. One-off human artefacts that happen to affect our lives, happy or unhappy accidents, can be safely excluded from the basic package of human extended parts. A haphazard scattering of trash might attract prey food-stuffs, but scattered trash should not be counted among our parts. Although, reliability can emerge from initially unreliable or unlikely accidents. We can learn from unreliable mistakes and highly-advantageous artefactual accidents can scaffold their subsequent locking in. One narrative for the origin of human-canine symbiosis involves an initial step where dog ancestors were adventitious scavengers of our (unreliable stream of) wastes and our subsequent "exaptation" of their predatory alerts. Today, the hunting dog is a symbol of functional reliability and fidelity, and arguably has become more a part of our groups over time.

Reciprocal causation (3) is perhaps the most contentious. Depending on one's choice of function concept, different human and even animal artefacts will thereby be excluded. I intend a wider reading, and think the essentials of CR functionality should count towards extended parthood. In the case of bodily parts, many are merely CR functional without thereby being non-parts. Perhaps the appendix now has no SE or etiological function, but this doesn't matter to its status as a part. Some stretches of human DNA seem to have no *adaptive* biochemical activities and no observable effects of their deletion—no fitness effects. Nonetheless they are copied, they may mutate to some effect, and at very least they consume some of the available energy resources for their upkeep in the cell—they are CR functional and they are indubitably parts of us. What is more concerning is the idea of parts that have no reciprocal effect whatever. E.g., many animals reliably produce certain markings in their surroundings—tracks or traces—yet these are ordinarily not even CR functional. Every organism has some reliable causal chains extending outside of them, but not all of these circle round and affect their origin. Objects that arise as mere effects of organisms are not their extended parts.

Overall I have argued that extended parthood includes all extended phenotypes and some additional human and non-human animal artefacts, but is not parthood run amok, since it does exclude some intuitive non-parts.

7.4.2 Extended Part-Whole Classification

Parts of organisms are subject to the part-whole based classification spanning levels of biological organization. Though not Dawkins' intention, his EPT clearly does have implications for classification. Consider first the standard, non-extended view of organisms and artefacts,

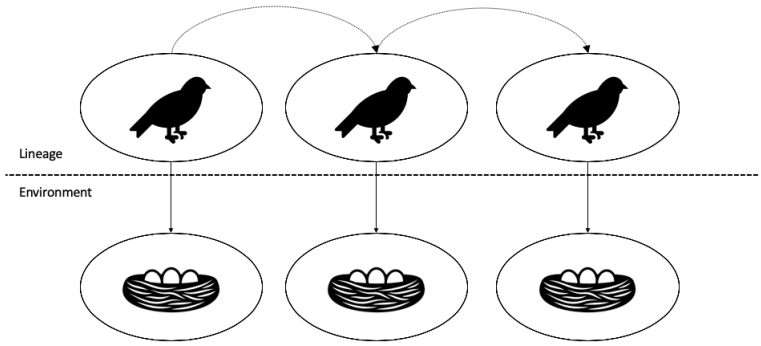


Fig. 7.9 The non-extended view of animal artefacts at the level of lineages. Lineages are of birds, their artefacts remain outside the lineage, since outside the body of the bird.

where there is a sharp divide between all animal artefacts and the organisms that produce them (Fig.7.8). Plainly, since on this view artefacts are not parts of organisms, they do not thereby have any place within the same part-whole classification. This makes them unclassifiable on the mereological account of classification. Notice further that this exclusion of extended parts at the organism level propagates upward through the classification at higher levels—artefacts are thereby not parts of lineages either (Fig.7.9).

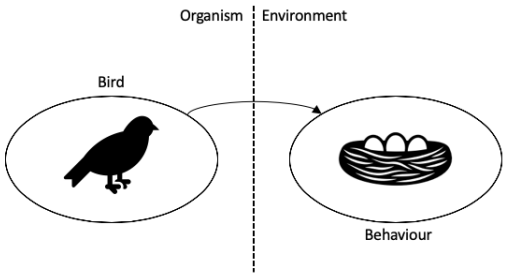


Fig. 7.8 The non-extended relationship between organisms and their artefacts. Artefacts are not parts of the organism, but parts of the environment.

The EPT changes how we think of the part-whole relationship between some artefacts and organisms. I think it best to treat the EPT as advocating a somewhat *more inclusive whole*, containing *both* artefacts and the "conventional" (or "bodily") phenotype (Fig.7.10). On this view, (some) artefacts are parts of biological-wholes and thereby subject to the part-whole classification scheme spanning levels of the biological hierarchy. Notice also that this inclusion of extended parts at the organism level propagates up to higher levels, as in Fig 7.11.

What I hope the account of extended parts adds to the EPT is just that more artefacts are extended parts than are extended phenotypes. Graphically, this view is captured by Figure 7.12. As with the EPT, this parthood relationship propagates upwards through the classificatory hierarchy. Since artefacts are parts of more inclusive biological wholes they are thereby subject to part-whole classification (Fig.7.13).

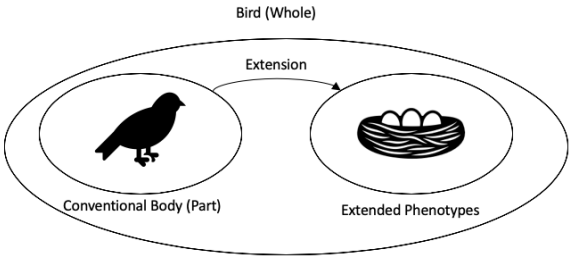


Fig. 7.10 The EPT view of (some) animal artefacts.

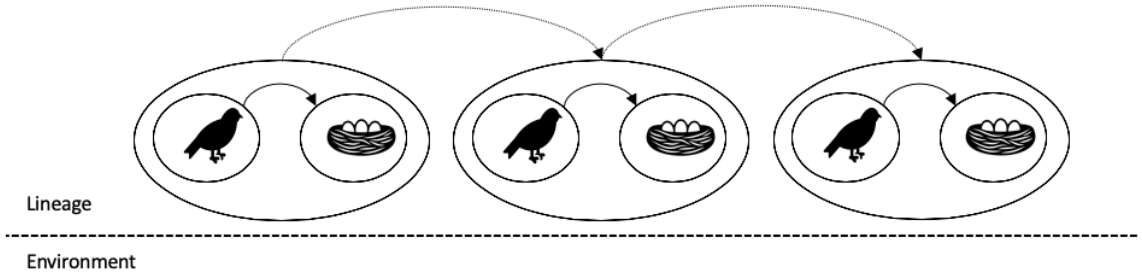


Fig. 7.11 The EPT view at the level of extended-lineages.

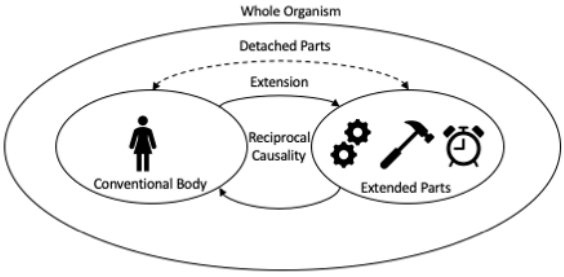


Fig. 7.12 A generic view of artefacts as extended parts. NB: human bodies and human artefacts are chosen for example and, since more general, the view should apply equally well to usual non-human artefacts.

When we see machines as extended parts of us, as human-parts, they earn a place in our classification of ourselves. As with the rejection of phenetic classification generally, it is no matter that machines are not sufficiently similar to our other parts to merit an hierarchical classification on the basis of overlapping sets of similarities. Similarity can be a bad indicator of history; a nest is not much like a mating dance and is much like certain happenstance amalgamations of leaf-litter, but the former are phylogenetically related in a deep sense that the latter are not. Machines do not usually much resemble organs, but are much more like

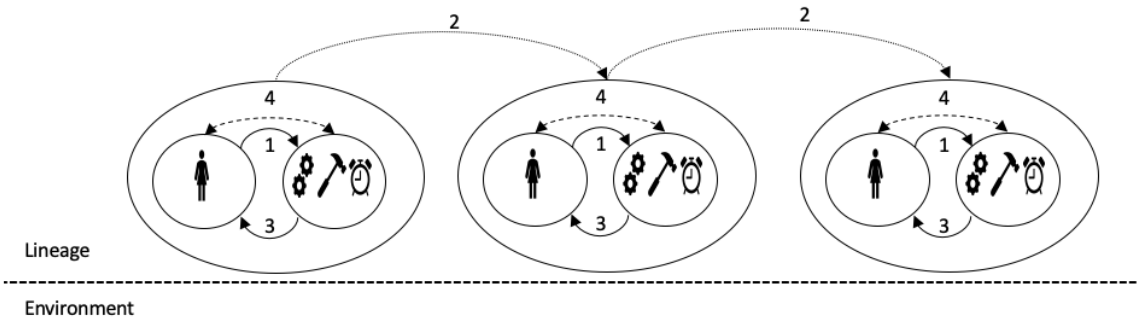


Fig. 7.13 The extended-parthood relationship at the level of lineages. Depicting (1) *causality*, (2) *reliability*, (3) *reciprocal causality* and (4) *detachment* is presented via the non-overlapping circles.

them than the are like mineral deposits, on phylogenetic grounds. Godfrey-Smith (2014 p.117) asks us to imagine zooming in on different parts of the "tree" of life, and finding different lower-level relationships in different branches. If we were to zoom in on history of life, including all the extended parts, machines would appear within the animal kingdom.[9]

7.5 Extendedness is not an MOD

Recalling the arguments of Ch.2, some organisms are themselves artefacts of other organisms, such as cows and other forms of animal-animal domestication (see Brooker and Feeney 2019). This suggests immediately that there could be overlap between the extended parts and organisms. I have argued that a theory of extended parthood, readily generalized from the EPT, broadens our biological mereology and thereby provides a classification of human and non-human animal-artefacts. If one is going to understand animal-artefacts within biology, then introducing the notion of an 'extended part' suffices. To conclude, I argue that this is not all it suffices for. This section argues that extended parthood applies to some *organisms themselves*—particularly in social animals. If one is going to understand artefact-animals, 'extended parthood' suffices there too. This shows that extended parthood not only resolves another incongruity in our classifications, but that it does not amount to an

[9]It might be complained that extended lineages are not very lineage-like, or at least that their "trees" are not very tree-like. This was essentially the problem dealt with in Ch.6. To my mind, that is more an issue with the relative importance of tree-likeness in evolutionary history, and indeed, the "tree" of non-extended lineages is already not very tree-like, given the preponderance of HGT. Moreover, if we see the replicator framework as a generalization of the notion of genes and their vehicles, and the ERF as a further generalization thereof to include non-genetic replicators, then noting the non-treelikeness of artefact histories, extended-lineages, is just to note that there is a process, akin to HGT, at the level of generality of ERF. That is, there is "Horizontal Replicator Transfer", but some of those replicators are not genes.

MOD: some machines are extended parts of organisms and some organisms are extended parts of organisms as well.

The *other* sort of biological phenomena that does not readily or intuitively fit within our part-whole classifications are eusocial relationships. As well as the animal-artefacts produced by social animals—bees, termites, social shrimp, ants, as well as beavers, primates, etc.—some of the animals themselves occupy a questionable position in our classifications. In particular, sterile worker animals and slave animals seem simultaneously to be organisms in their display of organization (they are free-floating amalgamations of cells that are organized into organs and organ-systems) while also being functional parts of larger organizations. For those sterile workers, their parthood relationships with lineages are doubly unusual: they are only ever at the tips of lineages, evolutionary "dead-ends"—like somatic cells—while they are nonetheless essential components for the perpetuation of "lineages" of the larger organizations. Referring to Fig.7.1, this places social animals in a position similar to that of artefacts—as "artefact-animals" instead of animal-artefacts—where their place in the part-whole classification of biology is uncertain.

Problems associated with understanding sociality at the level of lineages are old. The conceptual jump to considering selection of colony-level individuals was so great that it left Darwin in doubt (Herbers 2009). For Darwin, neuter insects were a "special difficulty, which at first appeared... insuperable, and actually fatal to my whole theory" (Darwin 2003 [1872]).[10] Perhaps unintuitively, I think we gain some traction by asking about the artefacts of social organisms: What is the appropriate mereology in cases of artefacts arising from social interactions?

The key animals cases involved are complex artefacts, like buildings, ant-hills, termite mounds and beaver dams. All seem to require some kind of collective or group-level contribution to extension, making it difficult to identify the whole that such structures are parts of. This is undeniably a key problem with considering human artefacts as extensions: the causal processes leading to human artefacts are much more complex, roundabout and amorphous than simpler causes of animal artefacts with a clear bearer—such as the web of a solitary spider. Dawkins, who is not shy of machine analogies, offered one on the topic of those extended phenotypes which appear to be extensions of more than one individual. After discussing the case of beaver habitats, which are often maintained by a family of beavers, he turns his attention to a case that is "worse": termite mounds. Whether this case is "worse"

[10] Whether the higher levels of organization apparently present in social animals necessitates a novel *sort* of organism concept, *superorganizmality,* has received substantial attention (Queller and Strassmann 1998; 2009). Another solution to the problem of neuter insects is provided in the theory of *inclusive fitness* (see debate between Abbot et al. 2011 and Nowak et al. 2010) where focus on lineages is shifted from considering the effects of an organism on its own direct descendents to those on its close relatives—resolving some of the strangeness of how non-reproductiveness of sterile insects could be adaptive.

depends, we will see, on how we conceive of the mereology of termite colonies. Dawkins says,

> Worse [than the case of beavers], consider the mound built by a colony of termites... It is built by perhaps a million termites, separated by time and into cohorts, like medieval masons who could work a lifetime on one cathedral and never meet their colleagues who would complete it. A partisan of the individual of the unit of selection might pardonably ask exactly whose extended phenotype the termite mound is supposed to be.—Dawkins (1982) p.200

Analogies with the structure of cathedrals have a special place in the hearts of philosophers of biology.[11] The problem for Dawkins is whether an organism, the worker termite, can have an extended phenotype that is the result of collective, group-level activities. His answer to this hypothetical objection is a *tu quoque*: that "exactly the same problem" apparently arises when multiple genes contribute to an organism's phenotype. That might be approximately right; there is collective causal contribution in both cases. The "partisan of the individual as the unit of section" will still object, however—Dawkins' retort does *not* pin the mound on a particular individual termite, and even a "superorganism" is a non-canonical individual at best.

Nonetheless, there is a direct answer to this hypothetical objection that even the partisan should be amenable to: *the mound is the extended phenotype of the queen, and so are the worker bodies*. Recall Dawkins' argumentative strategy: extended phenotypes are justified as a little step from ordinary phenotypes, since both involve "long, ramified and indirect causal chains". Dawkins seems to drop this approach precisely when continuing with it would be doubly beneficial. By abandoning this strategy, Dawkins is forced into embroiling himself in issues about (whole) colony-organism analogies and into questions about the proximal vs. distal effects of genes (p.199). In any case, however distal, workers p do satisfy condition (1) of the account of extended parts for their queen O—they are caused by her. They are also (2) reliably re-caused by her, as reliably as her "reproductive" system allows. They are also (3) functional for their queen, indeed even "sacrificing" their own evolutionary interests for hers. And (4) they are detached from her body after "birth".

We might imagine a more strategically stubborn Dawkins arguing that, just as bodily phenotypes are caused by long and ramified causal chains and some artefacts are caused by causal chains involving behaviour, others are cause by long causal chains involving *reproductive* behavior. One way evolution can lead to an organism constructing its house is by organisms that can build their own houses, under the power of their own bodies,

[11] Also see Odling-Smee and Turner (2011).

i.e. by creating organisms with straightforward extended phenotypes. Another is to create organisms that can get other bodies to produce the houses for them. One way to do *that* is by producing those bodies via a "reproductive" system, now coopted into worker manufacture.[12] There is also a very plausible origin for this sort of strategy since organisms typically come pre-equipped with the capacity to produce other bodies via the reproductive system. Provided the reproductive system can be bifunctional—making worker bodies and ordinary reproduction—it can be put to work extending the phenotype.

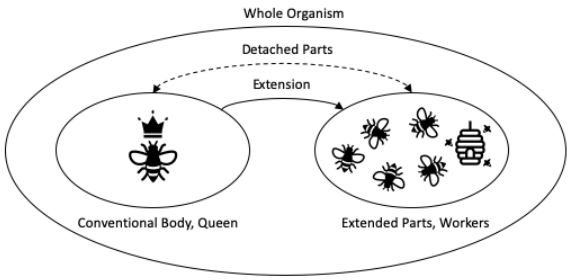


Fig. 7.14 Worker bees and hive as extended parts.

On this reading, the queen has many organs for mound construction that lie outside her conventional body and that have most of the trimmings of bodies themselves—legs, a nervous system, etc. In stubbornly extending Dawkins we begin to sound like Canguilhem. In the terminology of Canguilhem, individual worker termites are organs. I think these viewpoints say approximately the same thing: these interpretations of the caste systems provide further support to the idea that organisms can be extensions of organisms.

I conclude with the idea that the similarities between beavers and cathedral masons justifies thinking of complex artefacts and extensions of human groups.

On the basis of the analogy with cathedrals, one might incorrectly imagine that Dawkins was friendly to the idea that complex human artefacts are extended phenotypes. He is not. Though this is for good reason: cathedrals are not the products of a central genetic control by a sole reproductive member who is thereby the beneficiary of reciprocal effects of the cathedral.[13] If cathedrals are analogous to an animal artefact, termite mounds are not it. Dawkins takes the termite case to be "worse" than the beaver case because there are more termites in a mound than beavers in a family. This quantitatively complicates the process of explaining the emergence of mounds as an extended phenotype. With termites, the "quantitative, mutually interacting, mutually modifying, effects on a shared phenotype, the mound" [ibid, p.201] are greater in number. But on the reading given above, that termites

[12]Moreover this latter strategy comes with certain advantages not possible when building one's own house, such as a distribution of simultaneous labor—some bodies may eat while others build. Of course, this sort of strategy is only advantageous to the queen, or stably so, under certain conditions required to suppress the individual interests of these other bodies, such as sterility and non-aggression, what Godfrey-Smith (2009) calls "de-darwinization" of parts.

[13]Excluding, perhaps, the case of hereditary monarchies.

are extended parts of their queen, the beaver case is qualitatively "worse". In the termite case, despite being greater in number, the ramified causal chains bringing about phenotype are actually of a simpler sort: they are tree-like and rooted in the queen. In the causal process producing mounds, we may begin with the genes of the queen (if we like), then their influence on the production of the queen's conventional phenotype, then to the production of her extended phenotype, first via a large number of ramifications producing the external parts that are her workers, then to the direct effects of workers on the building or maintenance of mounds.

Fig. 7.15 Beaver stamp produced in 1851. From CBC (Nov 2019): https://www.cbc.ca/news/canada/beaver-canada-symbols-1.3658037

Beavers do not have queens, nor a worker cast that is easily amenable to consideration as extended parts of any other species member. It is beavers that are much more like masons constructing a cathedral—even provided we do not think too much philosophical ice is cut by an analogy to literal monarchies and worker castes. Beavers are the worse case because they build their complex artefacts in a family structure without clear genetic or control-hierarchy. When considering the ramifications of causal chains leading to the production of beaver dams and nests, there is no root in any particular beaver, no sole genetic beneficiary. For beavers, the "quantitative, mutually interacting, mutually modifying, effects on a shared phenotype" are organized into a complex partially-reticulated network, and just the same can be said of masons building a cathedral and of nearly every other complex human artefact.

If human artefacts are difficult to incorporate into biological theory on account of the complexity of the causal structure by which they arise, then so too for beaver dams. We need to allow for extended parts even if the causal chain leading to that part does not originate entirely with one organism, but arises from a collection. Perhaps we should allow that a beaver lodge be an extended phenotype of a beaver even when that beaver has not yet contributed much to the lodge, because its living relatives, parents or more distant ancestors

contributed. This is one way to make extended parts apply to structures produced by family or guild-like organisms. Another is to allow for extended-parthood at all levels of organization. If we want to say "exactly *whose* extended phenotype the [shelter] is supposed to be", perhaps the beaver-family, more than the termite-colony, deserves that title. That is, we might consider guild- or family-level extended traits. Canguilhem seems to have something at this level in mind when he says that machines are "organs of the human species" (1947). Beaver lodges are organs of the beaver family, and a young pup yet unable to produce or maintain the lodge is granted higher-level extended parts simply for its family membership.

Happily, all this can be justified "internally", on the basis of the intent to save the analysis of common animal artefacts like beaver lodges. Extension is a good way to establish a relationship between organisms and artefacts. Nonetheless, it is not an MOD. Some machines, like clocks and cathedrals, are extensions, and so are some organisms, like cows and termites. This is an un-Dawkinsian conclusion, but the conceptual move to achieve it is essentially Dawkinsian: artefacts are extended parts since they are products of specific sorts of behaviour, come whatever complexity of the causal structure needed to produce them.

Fig. 7.16 A beaver lodge. Photo taken by me, in Glengary, Southern Ontario, 2018.

7.6 Conclusion

Seeing technology as an extension of organisms entails that technology is not exogenous after all—in particular, that cyborgs are not, ultimately, "exogenously extended". Technology is a biological product and thus an endogenous element of life. A proper classification of technology should be included within biological classification of artefacts. The foregoing is offered in favour of an account of biological mereology that allows for "extended parts". The considerations of § 7.2 pose a challenge within biological theory to account for structures (artefacts) and relationships (enslavement, sociality) that do not fit nicely into a simple hierarchy. The solution to this problem within biological mereology suggested here is to allow extended parthood into our analysis. This lets us keep the conventional biological hierarchy much as it was by separating out the class of extended parts. Moreover, a theory

of extended parthood sufficient to include animal artefacts also ends up including machines as biological parts. The tactic has been to show that, with minimal elaboration on and generalization of existing theories of 'extension', a theory of extended parthood could be crafted that relates machines to organisms in a biologically familiar way. This "machine-organism relationship" is the closest we can get to an MOD. However, if we were looking for an MOD, this is not it. Social organisms likewise fit the mould of extended parts, so both organisms and machines are sometimes extended parts of other organisms.

Chapter 8

Conclusion

8.1 Recapitulation of Results Against the MOD

Is all this a vindication of Cartesian mechanical philosophy? Consider again what Descartes says in his infamous, and often grossly misinterpreted (Riskin 2016), simile about organisms and clocks.

> I desire, I say, that you should consider that these functions follow in this machine [an organism] simply from the disposition of the organs and *wholly naturally as* the movements of a clock or other automaton follow from the disposition of its counterweights and wheels.—Descartes, *Treaties on Man*, p.169, my emphasis

This thesis does not vindicate this view; I do not defend any reduction of the movements of animal bodies to the dispositions of their organs and make no use of the similar reduction of the movements of automatons to the disposition of their parts. Perhaps Descartes reductionism to dispositions is correct, or at least correct insofar as he takes it to exclude any need for "vegetative or sensitive souls" in the explanation for the movements of the bodies of organisms.

Moreover, taking the simile alone—made by the comparison "wholly naturally as"—I think the view is obviously correct, perhaps even a truism given a broad reading of Nature. Something about Descartes' phrasing make it seem as though a strong claim is being made, something like an identity or direct analogy between "organs" and "counterweights and wheels", when that is not really at issue. The simile is symmetric, however its reversal does not seem to arouse the same suspicion. If Descartes had claimed that *the movements of a clock or other automaton follow from the disposition of its counterweights and wheels* wholly naturally as *the functions in this [organism] follow from the disposition of the organs*, I think we wouldn't blink. Other passages support this reading: "all things that are artificial are...

natural. Thus, for example, when a watch keeps time by means of the wheels of which it is made, that is no less natural for it than it is for a tree to produce its fruit." (Descartes, Principles of Philosophy, Part iv, § 203, trans. Hacking 1998). Surely, we believe that the functions of organisms follow *naturally* from the dispositions of their organs if they follow from the dispositions of their organs at all, and we have no suspicion that anything less or more *natural* happens in the movements of clocks.[1] We do, however, have myriad other beliefs about what differs between machines and organisms, some of which have been addressed in this thesis.

Teleology

In chapter 2, on teleology, I addressed the view that organisms and machines differed in the sort of teleology applicable to them. I began with the view of Nicholson (2013) that organisms possess intrinsic teleology while machines possess only extrinsic teleology. Nicholson's (2013) view of teleology is founded on a variety of neo-Kantianism, where attributions of teleology are made on the basis of the relationship between the organization of an entity and its relationship to ends. If the ends of an entity are something other than itself, then that relationship is extrinsic; if the ends of an entity are only itself, than that relationship is intrinsic. I then showed that there are many cases where this distinction failed to apply—where organisms failed to have intrinsic teleology or machines failed to have extrinsic teleology—and where the typology actually reversed the expected order of teleological judgements—where an organism had extrinsic teleology and machines had intrinsic teleology. I did this by showing that there are plausible cases where an organism had ends directed towards things other than itself and where machines had ends directed towards themselves.

Autonomy

Another way of seeing the counterexamples presented in chapters 2-3 is as a criticism that neo-Kantianism was not adopted with sufficient uniformity. Pushed to examine cases of machines with the right sort of intrinsic organizations, we can ascribe them intrinsic teleology as well. Chapter 3 in particular addresses what sort of organization is required of organisms, and shows again that there are counterexamples of both varieties. Maturana and Varela (1980) and Moreno and Mossio (2015) both, in similar fashion, argue for placing the highest

[1] Excluding, of course, if we have some uncharitably reductionist reading of "following naturally from the dispositions of the parts", where either naturalness or dispositions exclude reasonable forms of emergence. I set that aside here since it is addressed earlier.

importance on the concept of autonomous organisation for organisms, and a corresponding lack of autonomy or allonomy of machines. Chapter 3 then shows that, provided a sufficiently broad consideration of organization in both, that there are non-autonomous and allonomous organisms and that there are non-allonomous and autonomous machines. I show this by analysing autonomy into a number of self-directed capacities, and providing examples of machines and organisms possessing or lacking these to different extents. We can see this as a further application of neo-Kantianism to autonomous machines, and its inapplicability to some allonomous organism).

Processualism

Chapter 4, addressing the topic of processualism and distinctions provided therein, is the first to move from teleology to ontology. Drawing largely on the work of Dupré (2013-2018) and Dupré and Nicholson (2018), I identified four key features of interest for process philosophers in their analysis of organisms: (1) how process is constitutive of organisms, (2) how stabilization of organisms is hierarchical, (3) the kind of material flux that organisms undergo, and (4) the role of "top-down" stabilization. In each case direct claims about distinctions with organisms were provided (by Dupré and/or Nicholson), and I showed in each case how either, (a) the distinction fails to account for counterexamples of either type or, (b) the difference with machines was actually unhelpful to the ultimate processualist goal of analysing organisms as processes.

Above I said that the neo-Kantian philosophers of biology were not Kantian enough; here I claim that the neo-Heraclitean process ontologists were not Hericlitean enough—when it comes to machines. On the one hand, acceptance of *panta* rhei (*everything* flows) leads inevitably to the conclusion that machines are a kind of process as well, so any *specifically processual* critique of the relationship between machines and organisms must fail—they may turn out to differ in some way, but not in being processes. On the other hand, Heraclitus' philosophy was partly processual and partly about the unity of opposites. The often misquoted river passage—"on those who enter *the same* rivers, *ever different* waters flow"—has to be seen in conjunction with others that unify opposites, such as "the way *up* and the way *down* is one and the same" (Burnet 1908). Even setting aside counterexamples of contemporary machines with key processual features (Ch.4.3), a neo-Heracliteanism accounting for both of these philosophies might have said: for machines to operate at a constant, ever different gears must turn. Insofar as organisms and machines are seeming opposites, Heraclitus would likely have sought their unification.

Mechanism

Chapter 5 continues with metaphysical issues, addressing distinctions with machines that appear within debates about the role of (new) mechanism in analysis of organisms[2]. One distinction, advocated similarly by Skillings (2015) and Godfrey-Smith (2016), points to the "regularity" of mechanisms (and machines) as distinguishing them from the "stochasticity" of organisms. I argued against this on two fronts: firstly it is a terminological error to interpret 'regularity' as 'non-stochasticity'— both in the mechanisms of new mechanism and machines—since 'regularity' is a Humean property of connectedness or gaplessness in explanations (Machamer et al. 2000 p.3, see § 5.1) and not equivalent to predictability. There are indeed many mechanisms in molecular biology that are stochastic in the sense Skillings and Godfrey-Smith find in organisms, and moreover there are plenty of machines that involve parts with stochastic effects.

The second mechanistic distinction—offered as a criticism of new mechanism by Godfrey-Smith (2016) and taken as an argument for new mechanism by Austin (2016)—is the role of holism. On Austin's (2016) view, machines are not appropriate models for the analysis of organisms because their effects do not arise holistically from their parts (as do the functions of organisms), while new-mechanisms are appropriate because they do allow for such holistic phenomena. Here I return to counterexamples of both types: first showing that there are features of organisms that do not seem to require an holistic analysis, and second showing that there are holistic machines.

These criticisms and advocations do get at something: we should not neglect stochasticity and holism from our analysis from the beginning, insofar as these features are important for explaining what organisms do. I would add to this that we should not neglect these features of machines, or mechanisms, either. Though I do not think they have been neglected in either case. What we end up with then in not a distinction between machines and organisms, but only a mild Cartesian-revisionism: Descartes spoke too soon when we said the following (immediately after referring to the explanation of movement by counterweights and wheels above),

> ...To explain these functions, then, it is not necessary to conceive of any vegetative or sensitive soul, *or any other principle of movement* or life. —Descartes, Treaties on Man, p.169, my emphasis

He should have allowed that machines might have dispositions that are more stochastic or holistic—more like lottery number generators or evolved microchips than counterweights

[2] § 5.1 contains an overview of the philosophy of new-mechanism and § 5.3 examines the older mechanical philosophy of Descartes

and wheels—and for the functions of organisms to be explainable by other, comparatively more statistical, principles of movement.

Evolution

Chapter 6 addressed the role of evolution in defining an MOD. The success of evolutionary theory has encouraged its application to a number of alternative cases. Its application to cultural change, technological change and specifically change in artefacts has been well studied. One common line of argument (Fracchia and Lewontin 2005) is that the theoretical assumptions required for evolution by natural selection are not appropriate to artefacts (machines) and another, more moderate, is that the assumptions do apply to artefacts but do not provide significant theoretical gains (Lewens 2004). In this chapter I begin by arguing that, even if the theoretical gains of an evolutionary account of technology are few, or only at high-levels of abstraction, that an evolutionary account explains the change of artefacts in part is enough to make it an essential component of our account of technology change.

I then rehearse a now standard line, that some artefacts indeed satisfy the Lewontin conditions for Evolution by Natural Selection (ENS)—phenotypic variation, differential fitness, and heritability—but that the cases where they interestingly satisfy them are few (largely restricted to digital cases and laboratory engineering experiments). I then argue that there are two other accounts of ENS that apply more interestingly to machines, and to more of them. These are, firstly, a generalization of the extended replicator framework of Sterelny et al. (1996), and a generalization of the conception of progenote evolution due to Woese (Woese 1998; Woese and Fox 1997). In sum, the chapter argues that there are three senses or modes of biological evolution applicable to machines.

Classification / Parthood

Chapter 7 addressed hierarchical biological classification. I take the best theory of biological classification to be coextensive with a theory of biological parthood, so the chapter focuses on considerations of parthood in organisms and machines. Specifically, I argue that machines *qua* human *artefacts* are not in a fundamentally different position from animal artefacts, both being plausibly captured by a modification of Dawkins (1882) theory of extended phenotypes. Moreover, it is not just that machines are much like the artefacts of organisms, I also argue that some social organisms themselves are plausibly considered as extended phenotypes. In service of this unification of diverse cases, I present a theory of extended parthood that is more general than those appearing in Dawkins (1982) or Sterelny et al. (1996).

Reconsidering the neo-Kantian underpinnings of chapters 2-3, the teleology and organization that Kant thought characteristic of organisms, as opposed to machines, was "grounded...on the *internal* possibility of the object" (Kant, *Judgement XII*, my emphasis). One way to look at the arguments of chapter 7 is another kind of challenge to this idea, showing that function and organization depend integrally on the extended possibilities of objects.[3] Once we account for organs like webs and nests, the grounding on internal vs. external possibilities of objects—that was supposed to neatly distinguish machines from organisms—becomes even more thoroughly blurred.

8.2 How to Think About Machines and Organisms as Kinds

Might we salvage something like an MOD by combining MODs discussed previously? And how would this be done? Biologists and their philosophers are no strangers to multi-factorial distinctions where no single property serves to demarcate a fundamental (natural) kind distinction. One way of going about this, common in science and growing in popularity among naturalist philosophers, is to produce a *space* of orthogonal distinctions, then plotting examples and their apparent counterexamples into the space. If some way of opposing axes in such a space succeeds in grouping cases into clusters, then we have succeeded in defining a multi-factorial distinction or "cluster kind". Moreover, when paired with a notion of some "homeostatic" process responsible for maintaining clustering, one obtains a homeostatic property cluster kind (Boyd 2010; Dieguez 2013).

An example of a clustering approach was used in Ch.2 (§ 2.6, Fig.2.1) to oppose two dimensions of analysis of teleological properties (rational-natural, and intrinsic-extrinsic). Another now famous example is Godfrey-Smith's (2009, chapter 3, Figure 3.1) "cube" relating three variable axes of interest for defining Darwinian populations (fidelity of inheritance, dependence of fitness on intrinsic differences, and smoothness of the fitness landscape). And another example is the graph by Queller and Strassmann (2009, Figures 1-3) plotting axes relevant to social organization (cooperation, conflict), where "organisms" cluster together in a region of high cooperation and low conflict, while societies cluster in the adjacent quadrant of high cooperation and high conflict.

For the spectrums of Intrinsic-Extrinsic teleology, Allonomy-Autonomy, and Natural-Rational history discussed in Ch.2-3, the following cube could be given. Ideally, if these axes defined an MOD, there would be a distinctly defined cluster of organisms (left-bottom-front)

[3]Put suggestively, if Darwin was Kant's "Newton of a blade of glass", then Dawkins was the Kant of spider-webs.

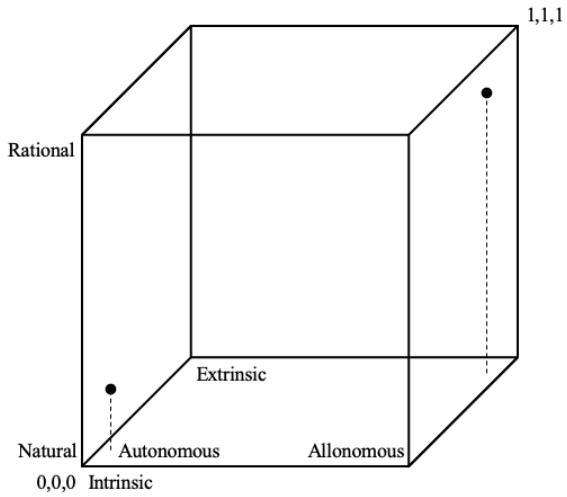


Fig. 8.1 Conceptual cube relating axes of Autonomy (x-axis), History (y-axis) and Teleology (z-axis).

and a cluster of machines (right-top-left). However the arguments of Ch.2-3 show that, though there may be clusters, they are not distinct.

An immediate problem arises when more than three axes seem relevant to clustering cases of interest: a cube only has three axes. If, for example, we wanted to see how Darwinian populations cluster when accounting for social organization, if we wanted to fit Godfrey-Smith's (2009) cube together with Queller and Strassmann's (2009) graph, we would have five axes. Granted we can plot a maximum of three at once, we would need ten cubes. What began as a conceptual aid has quickly turned into a conceptual difficulty of its own. If we want to think of something like a class or cluster based MOD, given the number of *prima facie* relevant dimensions to this analysis, we will need to think outside of the Godfrey-Smith cube and deploy a visual analogy for the conceptual landscape that is capable of expressing more multi-variate notions.[4] However we choose to represent the relationships between machines and organisms, the conceptual point remains: it will depend in part on what axes we find relevant and how we oppose these to one another. This thesis has shown that no matter one's choice from among the major competing analyses of machines and organisms, the clustering obtained will not be sharp and in each case there will be outliers—machines clustering better with organisms along one axis or another.

[4]I will not here defend any particular representation, since there are many available—examples include parallel coordinates where points in an n-dimensional space are depicted as lines with n-possible turning points (e.g., a radar plot), and techniques of dimensional reduction where similar dimensions are collapsed (e.g., principle-component analyses).

Might the MOD be found at this (higher) level of analysis of kinds of distinctions between clusters themselves? Are organisms and machines, not different kinds, different clusters, but different kinds of kinds? One way we might run this high-level MOD is to say that organisms (perhaps Life generally) is a homeostatic property cluster kind, while machines are not. That is, that the kindhood of organisms is maintained in homeostasis by some property or characteristic process applicable to them, while machines are not so maintained. In Ch.3-4, I argued that, with respect to properties of self-maintenance and bottom-up processes of physiological homeostasis, there was immense diversity among both organisms and machines. If it is bottom-up self-maintenance that is at issue, then there is indeed a difference between the homeostatic kindhood of machines and organisms, but it is a difference of degree—one sometimes violated—not a difference in kind of kinds.

Another potential homeostatic property is top-down, evolutionary stabilization. Organisms can be subject to various forces of "stabilizing selection", and moreover sexual organisms maintain species boundaries and within-species "stasis", over the short term at least, by recombination and species specific mating. Perhaps these properties serve to maintain organisms as a homeostatic kind. Nonetheless, in Ch.6-7, I argued that the processes that maintain technologies at higher levels are in may ways genuine processes of selection, and this can be seen as a top-down form of stabilization of technologies. Insofar as machines cluster together well, part of this clustering is explained top-down by forms of technological stabilizing selection.

In the latter chapters I aimed for unification between our analyses of machines and organisms, and I conclude with one further speculation on that topic. Perhaps cyborgs in the widest possible sense—encompassing nesting birds and humans with tools, as hybrids, consisting of organisms and the artefacts that arise from them—are actually high-probability outcomes of these high-level processes of, initially biological, evolution. Perhaps the cases that trouble us most arise more or less "naturally" as a consequence of the evolution of extended homeostatic properties or processes in organisms. The evolution of capacities to extend one's homestatic processes outside the conventional body into artefacts, and the eventual stabilization of integrated lineages of these extended parts (*qua* replicators or phenotypes) does seem to have arisen many times in a great diversity of organisms. If so, perhaps the union of organisms and their artefacts is itself a homeostatic property cluster kind.

8.3 Conclusion

The sceptic of the value of analytic counterexamples might rightly point out that, for many purposes in science and technology, the sort of refutation of the MOD offered in this thesis will come to little practical effect. They might point out that we can, nonetheless, often make judgements about which biological theories apply to contentious cases, and we are most often, in most familiar cases of organisms and machines, unperplexed by their overall similarities or differences. Indeed, throughout this thesis I have required and largely presupposed that we can sometimes identify cases of machines and organisms. To provide the counterexamples, it was necessary to claim that something is prima facie a machine (or organism) and possessed some property that some MOD said it ought not to possess. Moreover, at some points in this thesis it was necessary to make claims, in full seriousness, that would, in commonsense discussion of machines and organisms, probably sound like sarcasm ("clocks are not much like ribosomes", p.106). To the claim that an unsophisticated distinction can often suffice, I can only wholeheartedly agree. This thesis does not defend any broad scope indiscernibility of machines and organisms, it is not postmodern boundary blurring run amok. I have only argued that the MODs used by philosophers of biology did not do the analytic work sought of them, and they often did not do much more work than would be done simply by noting these more commonsense similarities and differences between machines and organisms.

I have not argued against all of the MODs that there could be, nor all that there are, but I have argued against all of the MODs that appear in contemporary philosophy of biology. I have shown how each MOD fails to account for our knowledge of present biological and technological diversity. In each case, I present counterexamples to the distinctions between machines and organisms that are offered, and sometimes show how these examples affect the conclusions that were supposed to be drawn from the MOD. In each case I also argue that the diversity spanning between machines and organisms should more properly be considered as a spectrum, full of intermediate cases and occasionally reversals of the expected relationship between machines and organisms. Each MOD offered by contemporary philosophers gets at something interesting and comes with far-ranging conclusions about organisms and machines; none establishes a fundamental distinction.

References

1. Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A., Andersson, M., ... & Barton, N. (2011). Inclusive fitness theory and eusociality. Nature, 471(7339), E1-E4.

2. Adleman, L. M. (1988, August). An abstract theory of computer viruses. In Conference on the Theory and Application of Cryptography (pp. 354-374). Springer, New York, NY.

3. Austin, C. J. (2016). The ontology of organisms: Mechanistic modules or patterned processes?. Biology & Philosophy, 31(5), 639-662.

4. Bapteste, E., & Dupré, J. (2013). Towards a processual microbial ontology. *Biology & Philosophy,* 28(2), 379-404.

5. Barwich, A. S. (2014). A sense so rare: Measuring olfactory experiences and making a case for a process perspective on sensory perception. Biological Theory, 9(3), 258-268.

6. Bateson, P. (1978). Review: The selfish gene: By Richard Dawkins. Oxford University Press (1976).

7. Bechtel, W. (2011). Mechanism and biological explanation. Philosophy of science, 78(4), 533-557.

8. Bedau, M. A. (2003). Artificial life: organization, adaptation and complexity from the bottom up. Trends in cognitive sciences, 7(11), 505-512.

9. Bennett, J. A. (1986). The mechanics' philosophy and the mechanical philosophy. History of science, 24(1), 1-28.

10. Benveniste, R. E., & Todaro, G. J. (1974). Evolution of C-type viral genes: inheritance of exogenously acquired viral genes. Nature, 252(5483), 456-459.

11. Benveniste, R. E., Sherr, C. J., & Todaro, G. J. (1975). Evolution of type C viral genes: origin of feline leukemia virus. Science, 190(4217), 886-888.

12. Bouchard, F. (2008). Causal processes, fitness, and the differential persistence of lineages. Philosophy of Science, 75(5), 560-570.

13. Boyd, R. (2010). Homeostasis, higher taxa, and monophyly. Philosophy of Science, 77(5), 686-701.

14. Brooker, R. M., & Feeney, W. E. (2019). Animal domesticators. Current Biology, 29(22), R1168-R1169.

15. Brunet, T. D. P. (2016). Disciplinology: Representations of the Structure of Science and the Humanities (Masters dissertation, Dalhousie University).

16. Brunet, T. D. P., & Doolittle, W. F. (2018). The generality of Constructive Neutral Evolution. Biology & Philosophy, 33(1-2), 2.

17. Burnet, J. (1908). Early Greek Philosophy.

18. Burrows, M., & Sutton, G. (2013). Interacting gears synchronize propulsive leg movements in a jumping insect. science, 341(6151), 1254-1256.

19. Butler, S. (1863). Darwin among the machines. To the Editor of the Press, Christchurch, New Zealand, 13 June

20. Canguilhem, G. (1952). Machine et organisme. In: La connaissance de la vie. Hachette, Paris [quoted from the fourth edition Vrin, Paris, 1971].

21. Castellanos, M. C., Wilson, P., & Thomson, J. D. (2004). 'Anti-bee' and 'pro-bird' changes during the evolution of hummingbird pollination in Penstemon flowers. Journal of evolutionary biology, 17(4), 876-885.

22. Cavanaugh, C. M. (1994). Microbial symbiosis: patterns of diversity in the marine environment. American Zoologist, 34(1), 79-89.

23. Clark, A. (2001). Reasons, robots and the extended mind. *Mind & Language*, 16(2), 121-145.

24. Clark, A., & Chalmers, D. (1998). The extended mind. analysis, 58(1), 7-19.

25. Clynes M. E., & Kline N. S. (1960). Cyborgs and space. *Astronautics*, septembre.

26. Cohen, F. (1987). Computer viruses: theory and experiments. Computers & security, 6(1), 22-35.

27. Covello, P., & Gray, M. (1993). On the evolution of RNA editing. Trends in Genetics, 9(8), 265-268.

28. Craver, C. F., & Darden, L. (2013). In search of mechanisms: Discoveries across the life sciences. University of Chicago Press.

29. Crilly, N. (2013). Function propagation through nested systems. Design Studies, 34(2), 216-242.

30. Crilly, N. (2015). The proliferation of functions: Multiple systems playing multiple roles in multiple supersystems. AI EDAM, 29(1), 83-92.

31. Cummins, R. (1975) Functional Analysis. Journal of Philosophy 72: 741-765.

32. Dadachova, E., & Casadevall, A. (2008). Ionizing radiation: how fungi cope, adapt, and exploit with the help of melanin. Current opinion in microbiology, 11(6), 525-531.

33. Darwin, C. (2003 [1872]) The Origin of Species. Broadview Press, Carroll, J. editor.

34. Dawkins, R. (1982). The extended phenotype (Vol. 8). Oxford: Oxford University Press.

35. Dawkins, R. (2004). Extended phenotype: but not too extended. A reply to Laland, Turner and Jablonka. Biology and Philosophy, 19(3), 377-396.

36. Dawkins, R. (2006). The selfish gene. UK: Oxford University Press. (Originally published in 1976).

37. De Queiroz, K. (2007). Species concepts and species delimitation. Systematic biology, 56(6), 879-886.

38. De Sousa, R. (2017). Nature's Purposes and Mine. How Biology Shapes Philosophy: New Foundations for Naturalism, 141-160.

39. Deleuze, G.,& Guattari, F. (1988). A thousand plateaus: Capitalism and schizophrenia. Bloomsbury Publishing.

40. Descartes, R., & Hall, T. S. (1972). Treatise of man. Harvard University Press.

41. Diéguez, A. (2013). Life as a homeostatic property cluster. *Biological Theory*, 7(2), 180-186.

42. Doolittle, W. F. (2000). Uprooting the tree of life. Scientific American, 282(2), 90-95.

43. Doolittle, W. F. (2014). Natural selection through survival alone, and the possibility of Gaia. Biology & Philosophy, 29(3), 415-423.

44. Doolittle, W. F. (2017). Making the most of clade selection. *Philosophy of science*, 84(2), 275-295.

45. Doolittle, W. F., & Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. Proceedings of the National Academy of Sciences, 104(7), 2043-2049.

46. Doolittle, W. F., & Brown, J. R. (1994). Tempo, mode, the progenote, and the universal root. Proceedings of the National Academy of Sciences, 91(15), 6721-6728.

47. Doolittle, W. F., & Brunet, T. D. (2016). What is the tree of life?. PLoS genetics, 12(4).

48. Doolittle, W. F., & Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature, 284(5757), 601.

49. Dubilier, N., Bergin, C., & Lott, C. (2008). Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nature Reviews Microbiology, 6(10), 725.

50. Dupré, J. (1995). The disorder of things: Metaphysical foundations of the disunity of science. Harvard University Press.

51. Dupré, J. (2010). The polygenomic organism. *The Sociological Review*, 58, 19-31.

52. Dupré, J. (2013, June). John Dupré: Living Causes. In Aristotelian Society Supplementary Volume (Vol. 87, No. 1, pp. 19-37). Oxford, UK: Oxford University Press.

53. Dupré, J. (2017). The metaphysics of evolution. Interface focus, 7(5), 20160148.

54. Dupré, J. (2017). Book review: DM Walsh//organisms, agency and evolution. British Journal for the Philosophy of Science.

55. Dupré, J. (2018). Processes, Organisms, Kinds, and the Inevitability of Pluralism. Individuation, Process, and Scientific Practices, 21.

56. Dupré, J. A., & Nicholson, D. J. (2018). A manifesto for a processual philosophy of biology. (in *Everything Flows, D.J.N & J.A.D ed.*)

57. Dupré, J., & O'Malley, M. A. (2009). Varieties of living things: life at the intersection of lineage and metabolism.

58. Dutton, D. (2009). The art instinct: Beauty, pleasure, & human evolution. Oxford University Press, USA.

59. Ehrich, S., Behrens, D., Lebedeva, E., Ludwig, W., & Bock, E. (1995). A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, Nitrospira moscoviensis sp. nov. and its phylogenetic relationship. Archives of Microbiology, 164(1), 16-23

60. Ereshefsky, M., & Pedroso, M. (2013). Biological individuality: the case of biofilms. Biology & Philosophy, 28(2), 331-349.

61. Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. Robotics and autonomous systems, 42(3-4), 143-166.

62. Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. Virus research, 117(1), 5-16.

63. Fracchia, J., & Lewontin, R. C. (2005). The price of metaphor. History and Theory, 44(1), 14-29.

64. Fullmer, M. S., Soucy, S. M., & Gogarten, J. P. (2015). The pan-genome as a shared genomic resource: mutual cheating, cooperation and the black queen hypothesis. Frontiers in microbiology, 6, 728.

65. Garson, J. (2017). Against Organizational Functions. Philosophy of Science, 84(5), 1093-1103.

66. Ghiselin, M. T. (2005). Homology as a relation of correspondence between parts of individuals. Theory in Biosciences, 124(2), 91-103.

67. Glennan, S. S. (1996). Mechanisms and the nature of causation. Erkenntnis, 44(1), 49-71.

68. Godfrey-Smith, P. (1993). Functions: Consensus without unity. *Pacific Philosophical Quarterly*, 74(3), 196-208.

69. Godfrey-Smith, P. (2001). Three kinds of adaptationism. Adaptationism and optimality, 122.

70. Godfrey-Smith, P. (2009). Darwinian populations and natural selection. Oxford University Press.

71. Godfrey-Smith, P. (2011). The evolution of the individual. Lakatos Award Lecture, LSE.

72. Godfrey-Smith, P. (2012). Darwinism and cultural change. *Philosophical Transactions of the Royal Society B*: Biological Sciences, 367(1599), 2160-2170.

73. Godfrey-Smith, P. (2013). *Philosophy of biology*. Princeton University Press.

74. Godfrey-Smith, P. (2015). Reproduction, symbiosis, and the eukaryotic cell. *Proceedings of the National Academy of Sciences*, 201421378.

75. Godfrey-Smith, P. (2016). Mind, matter, and metabolism. *The Journal of Philosophy*, 113(10), 481-506.

76. Gogarten, J. P., & Deamer, D. (2016). Is LUCA a thermophilic progenote?. Nature microbiology, 1(12), 1-2.

77. Goldsmith, T. C. (2012). On the programmed / non-programmed aging controversy. Biochemistry (Moscow), 77(7), 729-732.

78. Gorham, G. (1994). Mind-body dualism and the Harvey-Descartes controversy. Journal of the History of Ideas, 55(2), 211-234.

79. Gould, S. J. (1988). An urchin in the storm: Essays about books and ideas. WW Norton & Company.

80. Gray, R. D., Greenhill, S. J., & Ross, R. M. (2007). The pleasures and perils of Darwinizing culture (with phylogenies). Biological Theory, 2(4), 360-375.

81. Griffiths, P. E., & Gray, R. D. (2001). Darwinism and developmental systems. Cycles of contingency: Developmental systems and evolution, 195-218.

82. Griffiths, P. E., & Gray, R. D. (2005). Discussion: Three ways to misunderstand developmental systems theory. Biology and Philosophy, 20(2-3), 417-425.

83. Hacking, I. (1998). Canguilhem amid the cyborgs. Economy and Society, 27(2-3), 202-216.

84. Hager, M. D., Greil, P., Leyens, C., van der Zwaag, S., & Schubert, U. S. (2010). Self-healing materials. Advanced Materials, 22(47), 5424-5430.

85. Haraway, D. (1990). A manifesto for cyborgs: Science, technology, and socialist feminism in the 1980s. Feminism / postmodernism, 190-233.

86. Hehmeyer, J. (2019). Two potential evolutionary origins of the fruiting bodies of the dictyostelid slime moulds. Biological Reviews, 94(5), 1591-1604.

87. Herbers, J. M. (2009). Darwin's 'one special difficulty': celebrating Darwin 200. Biology letters, 5(2), 214-217.

88. Hirst, N. F. (2008). Towards a science of life as creative organisms. Cosmos and History: The Journal of Natural and Social Philosophy, 4(1-2), 78-98.

89. Hull, D. L. (1986, January). On human nature. In PSA: Proceedings of the biennial meeting of the philosophy of science association (Vol. 1986, No. 2, pp. 3-13). Philosophy of Science Association.

90. Hutter, M., Gehring, C., Jud, D., Lauber, A., Bellicoso, C. D., Tsounis, V., ... & Diethelm, R. (2016, October). Anymal-a highly mobile and dynamic quadrupedal robot. In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 38-44). IEEE.

91. Ieropoulos, I., Melhuish, C., Greenman, J., & Horsfield, I. (2005). EcoBot-II: An artificial agent with a natural metabolism. International Journal of Advanced Robotic Systems, 2(4), 31.

92. Iliopoulos, D., Adami, C., & Szor, P. (2011). Darwin inside the machines: Malware evolution and the consequences for computer security. arXiv preprint arXiv:1111.2503.

93. Jablonka E. (2004), From Replicators to Heritably Varying Phenotypic Traits: The Extended Phenotyper Revisited?, Biology and Phylosophy 19, 353-375.

94. Jablonka, E. (2007). The developmental construction of heredity. Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 49(8), 808-817.

95. Jablonski, D. (2008). Species selection: theory and data. Annual Review of Ecology, Evolution, and Systematics, 39, 501-524.

96. Jacob, F. (1977). Evolution and tinkering. Science, 196(4295), 1161-1166.

97. Jacob, F., & Bendall, D. S. (1983). Molecular tinkering in evolution. Evolution from molecules to men, 131, 144.

98. Jankovic, S., & Cirkovic, M. M. (2016). Evolvability is an evolved ability: the coding concept as the arch-unit of natural selection. Origins of Life and Evolution of Biospheres, 46(1), 67-79.

99. Jones, R., Haufe, P., Sells, E., Iravani, P., Olliver, V., Palmer, C., & Bowyer, A. (2011). RepRap:the replicating rapid prototyper. Robotica, 29(1), 177-191.

100. Kaczynski, T. J. (2016). Anti-tech Revolution: Why and how. Fitch & Madison Publishers.

101. Kant, I. (2000). Critique of the Power of Judgment. Cambridge University Press. (original)

102. Karpenko, Y. V., Redchitz, T. I., Zheltonozhsky, V. A., Dighton, J., & Zhdanova, N. N. (2006). Comparative responses of microscopic fungi to ionizing radiation and light. Folia microbiologica, 51(1), 45.

103. Kitcher, P. (1993). Function and design. Midwest studies in philosophy, 18, 379-397.

104. Kochanowski, K., Volkmer, B., Gerosa, L., van Rijsewijk, B. R. H., Schmidt, A., & Heinemann, M. (2013). Functioning of a metabolic flux sensor in Escherichia coli. Proceedings of the National Academy of Sciences, 110(3), 1130-1135.

105. Koonin, E. V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. Nature Reviews Microbiology, 1(2), 127.

106. Koonin, E. V. (2014). Carl Woese's vision of cellular evolution and the domains of life. RNA biology, 11(3), 197-204.

107. Koonin, E. V. (2015). The turbulent network dynamics of microbial evolution and the statistical tree of life. Journal of molecular evolution, 80(5-6), 244-250.

108. Koonin, E. V. (2016). Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. F1000Research, 5.

109. Koonin, E. V., & Wolf, Y. I. (2012). Evolution of microbes and viruses: a paradigm shift in evolutionary biology?. Frontiers in cellular and infection microbiology, 2, 119.

110. Laland, K. N. (2004). Extending the extended phenotype. Biology and Philosophy, 19(3), 313-325.

111. Lewens, T. (2006). Organisms and artifacts: Design in nature and elsewhere. MIT Press

112. Lewens, T. (2013). From bricolage to BioBricks™: Synthetic biology and rational design. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44(4), 641-648.

113. Lewens, T. (2015). Cultural evolution: conceptual challenges. OUP Oxford.

114. Lewens, T., (2020) Cultural Evolution, The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)

115. Lynch, M., & Marinov, G. K. (2017). Membranes, energetics, and evolution across the prokaryote-eukaryote divide. Elife, 6, e20437.

116. Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. Philosophy of science, 67(1), 1-25.

117. Mange, D., Goeke, M., Madon, D., Stauffer, A., Tempesti, G., & Durand, S. (1996). Embryonics: A new family of coarse-grained field-programmable gate array with self-repair and self-reproducing properties. In Towards evolvable hardware(pp. 197-220). Springer, Berlin, Heidelberg.

118. Mann, S., & Niedzviecki, H. (2001). Cyborg: Digital destiny and human possibility in the age of the wearable computer. Doubleday Canada.

119. Margulis, L., & Fester, R. (Eds.). (1991). Symbiosis as a source of evolutionary innovation: speciation and morphogenesis. Mit Press.

120. Mariscal, C., & Doolittle, W. F. (2017). Life and life only: a radical alternative to life definitionism. Synthese, 1-15.

121. Markvicka, E. J., Bartlett, M. D., Huang, X., & Majidi, C. (2018). An autonomously electrically self-healing liquid metal–elastomer composite for robust soft-matter robotics and electronics. *Nature materials*, 17(7), 618-624.

122. Maturana, H. R., & Varela, F. J. (1972). Autopoiesis and Cognition. The realization of the living. Dodrecht & Bosten, London: D. Reidel Publishing Company.

123. McAdams, H. H., & Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3), 814-819.

124. McShea, D. W., & Venit, E. P. (2000). What is a part?. The character concept in evolutionary biology, 259.

125. Melhuish, C., Ieropoulos, I., Greenman, J., & Horsfield, I. (2006). Energetically autonomous robots: Food for thought. Autonomous Robots, 21(3), 187-198.

126. Metzger, M. J., Reinisch, C., Sherry, J., & Goff, S. P. (2015). Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. Cell, 161(2), 255-263.

127. Militello, G., & Moreno, n (2018). Structural and organisational conditions for being a machine. Biology & Philosophy, 33(5-6), 35.

128. Millikan, R. G. (1984). Language, thought, and other biological categories: New foundations for realism. MIT press.

129. Moore, P. B. (2012). How should we think about the ribosome?. *Annual review of biophysics*, 41, 1-19.

130. Moreno, A., & Mossio, M. (2015). Biological autonomy. A Philo.

131. Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio, 3(2), e00036-12.

132. Mossio, M., & Moreno, A. (2010). Organisational closure in biological organisms. History and philosophy of the life sciences, 269-288.

133. Mossio, M., Saborido, C., & Moreno, A. (2009). An organizational account of biological functions. The British Journal for the Philosophy of Science, 60(4), 813-841.

134. Needham, J. (1926). New-Mechanism: or the Sceptical Biologist. Hibbert Journal, vol.25 1926-27

135. Needham, J. (1928). Organicism in biology. *Journal of Philosophical Studies*, 3(9), 29-40.

136. Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23), 8577-8582.

137. Nicholson, D. J. (2013). Organisms $\neq$ machines. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44(4), 669-678.

138. Nicholson, D. J. (2018). Reconceptualizing the organism: from complex machine to flowing stream. In *Everything Flows*, Dupré and NIcholson eds.

139. Nicholson, D. J. (2019). Is the cell really a machine?. Journal of theoretical biology, 477, 108-126.

140. Nicholson, D. J., & Dupré, J. (Eds.). (2018). Everything flows: towards a processual philosophy of biology. Oxford University Press.

141. Nietzsche, F. (1996). Nietzsche: Human, all too human: A book for free spirits. Cambridge University Press.

142. Nietzsche, F. (1997). Daybreak: Thoughts on the Prejudices of Morality, ed. Maudemarie Clark and Brian Leiter, trans. RJ Hollingdale. Cambridge: Cambridge University Press, 197, 197-198.

143. Novick, A., & Doolittle, W. F. (2020). Horizontal persistence and the complexity hypothesis. Biology & Philosophy, 35(1), 2.

144. Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. Nature, 466(7310), 1057-1062.

145. O'Malley, M. (2014). Philosophy of microbiology. Cambridge University Press.

146. O'Malley, M. A., & Dupré, J. (2005). Fundamental issues in systems biology. BioEssays, 27(12), 1270-1276.

147. O'Malley, M. A., & Dupré, J. (2007). Size doesn't matter: towards a more inclusive philosophy of biology. Biology & Philosophy, 22(2), 155-191.

148. Odling-Smee, F. J. (1988). Niche-constructing phenotypes. In H. C. Plotkin (Ed.), The role of behavior in evolution (p. 73–132). The MIT Press.

149. Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (1996). Niche construction. The American Naturalist, 147(4), 641-648.

150. Odling-Smee, J., & Turner, J. S. (2011). Niche construction theory and human architecture. Biological Theory, 6(3), 283-289.

151. Ohno, S. (1972). So much 'junk' DNA in our genome. In Evolution of Genetic Systems, Brookhaven Symp. Biol. (pp. 366-370).

152. Okasha, S. (2003). Does the concept of clade selection make sense?. Philosophy of Science, 70(4), 739-751.

153. Oparin, A. I. 1957. The Origin of Life on Earth. New York: Academic. 495 pp.

154. Orgel, L. E., & Crick, F. H. (1980). Selfish DNA: the ultimate parasite. Nature, 284(5757), 604.

155. Oyama, S., Griffiths, P. E., & Gray, R. D. (2001). Introduction: What is developmental systems theory. Cycles of contingency: Developmental systems and evolution, 1-11.

156. Pande, S., & Kost, C. (2017). Bacterial unculturability and the formation of intercellular metabolic networks. Trends in microbiology, 25(5), 349-361.

157. Payne, S. H., & Loomis, W. F. (2006). Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. Eukaryotic cell, 5(2), 272-276.

158. Penrose, L. S. (1951). Measurement of pleiotropic effects in phenylketonuria. Originally: Annals of Eugenics, 16(1), 134-141. Now appears in Annals of Human Genetics

159. Pfeifer, R., Lungarella, M., & Iida, F. (2007). Self-organization, embodiment, and biologically inspired robotics. science, 318(5853), 1088-1093.

160. Pham, V. H., & Kim, J. (2012). Cultivation of unculturable soil bacteria. TRENDS in Biotechnology, 30(9), 475-484.

161. Pollan, M. (2001). The botany of desire: A plant's-eye view of the world. Random house trade paperbacks.

162. Quammen, D. (2018). The tangled tree: a radical new history of life. Simon and Schuster.

163. Queller, D. C., & Strassmann, J. E. (1998). Kin selection and social insects. Bioscience, 48(3), 165-175.

164. Queller, D. C., & Strassmann, J. E. (2009). Beyond society: the evolution of organismality. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1533), 3143-3155.

165. Raibert, M., Blankespoor, K., Nelson, G., & Playter, R. (2008). Bigdog, the rough-terrain quadruped robot. IFAC Proceedings Volumes, 41(2), 10822-10825.

166. Rapp, F. (1981). Analytical Philosophy of Technology. Boston Studies in the Philosophy of Science New York, NY, 63, 1-199.

167. Rescher N. 2004 Process philosophy, The Stanford encyclopedia of philosophy (Spring 2004 Edition) (ed. Edward N Zalta)

168. Rescher, N. (1996). Process metaphysics: An introduction to process philosophy. Suny Press.

169. Richardson, A. O., & Palmer, J. D. (2007). Horizontal gene transfer in plants. Journal of experimental botany, 58(1), 1-9.

170. Rose, M. R., & Doolittle, W. F. (1983). Molecular biological mechanisms of speciation. Science, 220(4593), 157-162.

171. Rubenstein, M., Cornejo, A., & Nagpal, R. (2014). Programmable self-assembly in a thousand-robot swarm. Science, 345(6198), 795-799.

172. Sathe, S., Kaushik, S., Lalremruata, A., Aggarwal, R. K., Cavender, J. C., & Nanjundiah, V. (2010). Genetic heterogeneity in wild isolates of cellular slime mold social groups. *Microbial ecology*, 60(1), 137-148.

173. Scharf, C., Virgo, N., Cleaves, H. J., Aono, M., Aubert-Kato, N., Aydinoglu, A., ... & Brasser, R. (2015). A strategy for origins of life research.

174. Scriver, C. R., & Waters, P. J. (1999). Monogenic traits are not simple: lessons from phenylketonuria. Trends in Genetics, 15(7), 267-272.

175. Seibt, J. (2012). Process philosophy. Stanford Encyclopedia of Philosophy

176. Seibt, J. (2018). Ontological Tools for the Process Turn in Biology. Everything flows: Towards a Processual philosophy of biology, 113.

177. Sells, E.A. (2009). Towards a Self-Manufacturing Rapid Prototyping Machine', Ph.D., University of Bath.

178. Skillings, D. J. (2015). Mechanistic explanation of biological processes. *Philosophy of Science*, 82(5), 1139-1151.

179. Smith, J. M., & Szathmary, E. (1997). The major transitions in evolution. Oxford University Press.

180. Soble (2015). Robotica: A Robotic Dog's Mortality. New York Times, June 17, 2015

181. Stahl, W. R. (1965). Self-reproducing automata. Perspectives in biology and medicine, 8(3), 373-393.

182. Sterelny, K. (2001). Niche construction, developmental systems, and the extended replicator. Cycles of contingency: Developmental systems and evolution, 333-350.

183. Sterelny, K. (2005). Made by each other: organisms and their environment. Biology and Philosophy, 20(1), 21-36.

184. Sterelny, K. (2006). Memes revisited. The British Journal for the Philosophy of Science, 57(1), 145-165.

185. Sterelny, K., Smith, K. C., & Dickison, M. (1996). The extended replicator. Biology and Philosophy, 11(3), 377-403.

186. Suthakorn, J., Cushing, A. B., & Chirikjian, G. S. (2003, July). An autonomous self-replicating robotic system. In Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003) (Vol. 1, pp. 137-142). IEEE.

187. Thompson, A. (2002). Notes on Design through Artificial Evolution: Opportunities and Algorithms. In I.C. Parmee (Ed.), Adaptive Computing in Design and Manufacture (V). Springer: pp. 17–26.

188. Tomita, K., Murata, S., Kurokawa, H., Yoshida, E., & Kokaji, S. (1999). Self-assembly and self-repair method for a distributed mechanical system. IEEE Transactions on Robotics and Automation, 15(6), 1035-1045.

189. Van Melderen, L., & De Bast, M. S. (2009). Bacterial toxin-antitoxin systems: more than selfish entities?. PLoS genetics, 5(3).

190. Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. Biosystems, 5(4), 187-196.

191. Von Neumann, J., & Burks, A. W. (1966). Theory of self-reproducing automata. IEEE Transactions on Neural Networks, 5(1), 3-14.

192. Waddington, C. H. (1952). Selection of the genetic basis for an acquired character. *Nature*, 169(4302), 625-626.

193. Wade, N. (2005). Your body is younger than you think. New York Times.

194. Walsh, D. M. (2015). Organisms, agency, and evolution. *Cambridge University Press.*

195. Wernegreen, J. J. (2012). Endosymbiosis. Current Biology, 22(14), R555-R561.

196. Whitehead, A. N. (1929). Process and Reality, an Essay in Cosmology

197. Wiener, N. (2019 [1948]). Cybernetics or Control and Communication in the Animal and the Machine. MIT press.

198. Wilkins, J. S., & Bourrat, P. (2001). Replication and reproduction. Stanford Encyclopedia of Philosophy (accessed Jan 2021)

199. Williams, T. A., Cox, C. J., Foster, P. G., Szöllősi, G. J., & Embley, T. M. (2020). Phylogenomics provides robust support for a two-domains tree of life. Nature ecology & evolution, 4(1), 138-147.

200. Wilson, D. S. (1980). The natural selection of populations and communities. Benjamin/Cummings Pub. Co..

201. Woese, C. (1998). The universal ancestor. Proceedings of the national academy of Sciences, 95(12), 6854-6859.

202. Woese, C. R. (2002). On the evolution of cells. Proceedings of the National Academy of Sciences, 99(13), 8742-8747.

203. Woese, C. R. (2004). A new biology for a new century. Microbiol. Mol. Biol. Rev., 68(2), 173-186.

204. Woese, C. R., & Fox, G. E. (1977). The concept of cellular evolution. Journal of molecular evolution, 10(1), 1-6.

205. Wolf, Y. I., & Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. Bioessays, 35(9), 829-837.

206. Wright, L. (1973). Functions. The philosophical review, 82(2), 139-168.

207. Yim, M., Shen, W. M., Salemi, B., Rus, D., Moll, M., Lipson, H., ... & Chirikjian, G. S. (2007). Modular self-reconfigurable robot systems [grand challenges of robotics]. IEEE Robotics & Automation Magazine, 14(1), 43-52.

208. Zykov, V., Mytilinaios, E., Adams, B., & Lipson, H. (2005). Robotics: Self-reproducing machines. Nature, 435(7039), 163.